

STIC-Biotech/ChemLib

60964

From: Bunner, Bridget  
Sent: Monday, February 25, 2002 12:20 PM  
To: STIC-Biotech/ChemLib  
Subject: sequence search

Hi! I'd like a sequence search performed for case 09/755,017:

CRFE

1. the nucleic acid sequence of SEQ ID NO: 1
2. the nucleic acid sequence that encodes the amino acid sequence of SEQ ID NO: 2

thanks!

Bridget Bunner

Art Unit 1647  
CM1-10D12  
(703) 305-7148  
mailroom 10C01

Point of Contact:  
Toby Port  
Technical Info. Specialist  
CM1 ~~1E01~~ TEL: 308-3534

6104

	TYPE OF SEARCH:	VENDOR/COST(where applic.)
Searcher: _____	NA Sequences: <u>2</u>	STN: _____
Phone: _____	AA Sequences: _____	DIALOG: _____
Location: _____	Structures: _____	Questel/Orbit: _____
Date Picked Up: <u>2/25</u>	Bibliographic: _____	DRLink: _____
Date Completed: <u>2/26</u>	Litigation: _____	Lexis/Nexis: _____
Searcher Prep/Review: <u>10</u>	Full text: _____	Sequence Sys.: <u>es</u>
Clerical: _____	Patent Family: _____	WWW/Internet: _____
Online time: <u>10</u>	Other: _____	Other (specify): _____



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 03:36:01 (Search time 1214.49 Seconds  
(without alignments) 8334.794 Million cell updates/sec)

Title: US-09-755-017-1

Perfect score: 942  
Sequence: 1 atgaattggttaatgacag.....tcttcttaatcaagaataaa 942

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Pre-processing:	Minimum Match	0%
Post-processing:	Minimum Match	0%

Database :

```

1:  em_estfun:*
2:  em_esthnm:*
3:  em_estlin:*
4:  em_estlom:*
5:  em_estpl:*
6:  em_estpa:*
7:  em_estro:*
8:  em_estov:*
9:  em_hlc:*
10: gp_estli:*
11: gp_estr:*
12: gp_hlc:*
13: gp_gss:*
14: em_gss_fun:*
15: em_gss_hnm:*
16: em_gss_inv:*
17: em_gss_pli:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vtr:*
21: em_gss_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
C 1	337.4	35.8	479	13	A0077154	A0077154	CIT-HSP-2
C 2	203	21.5	523	10	A1604386	vv74c09..x	A1604386
C 3	201.8	21.4	692	12	A2090606	A2090606	RPCI-23-7
C 4	196.6	20.9	1394	12	AK017036	AK017036	Mus musc
C 5	196.6	20.9	3063	12	AK016560	AK016560	Mus musc
C 6	195.6	20.8	662	13	A2969227	A2969227	2M024J124
C 7	194.2	20.6	632	13	A2511623	A2511623	IM0356C17
C 8	185.4	19.7	428	11	N68399	N68399	za13b04..s1
C 9	184	19.5	326	13	A2694245	A2694245	AST-ZHBG3
C 10	180.4	19.2	788	13	BH113304	BH113304	RPCI-24-3
C 11	178.6	19.0	640	13	A2516219	A2516219	RPCI-11-3-3
C 12	175.6	18.6	588	13	A2413817	A2413817	IM0188F05

C	13	172	18.3	740	13	A2607393
C	14	164	17.4	448	13	A2271125
C	15	162.6	17.3	632	13	A2765752
C	16	159.6	16.9	680	13	A2709687
C	17	157.4	16.7	660	13	A2380178
C	18	156.2	16.6	675	13	A2396801
C	19	154.8	16.4	1501	12	AK015338
C	20	154.2	16.4	522	13	A2720043
C	21	153.8	16.3	646	13	A2638594
C	22	152.4	16.2	686	13	A2088625
C	23	152	16.1	534	13	A2312830
C	24	151.8	16.2	608	13	A2112932
C	25	151.8	16.1	797	13	A2909618
C	26	151	16.0	628	13	A2977433
C	27	150	15.9	488	13	A2373636
C	28	149.6	15.9	504	13	AO310400
C	29	149.4	15.9	632	13	A2392351
C	30	148.4	15.8	605	13	A2644411
C	31	148.2	15.8	680	13	A2235331
C	32	147.8	15.7	639	13	A2418543
C	33	147.8	15.7	796	11	BG197640
C	34	147.6	15.6	619	10	A1148854
C	35	147.4	15.6	532	13	A2519123
C	36	147.2	15.6	553	13	AO483125
C	37	147.2	15.6	614	13	AO453957
C	38	146.6	15.6	649	13	BH067870
C	39	146.4	15.5	534	13	A2664867
C	40	145.8	15.5	521	13	A2407823
C	41	145.8	15.5	678	13	BH067899
C	42	145.4	15.4	738	10	AI079550
C	43	145.4	15.4	762	11	BG193339
C	44	145.2	15.4	631	11	A2454523
C	45	145.2	15.4	797	13	A2685845
C	46	145.2	15.4	797	13	A2685845
C	47	145.2	15.4	797	13	A2685845
C	48	145.2	15.4	797	13	A2685845
C	49	145.2	15.4	797	13	A2685845
C	50	145.2	15.4	797	13	A2685845
C	51	145.2	15.4	797	13	A2685845
C	52	145.2	15.4	797	13	A2685845
C	53	145.2	15.4	797	13	A2685845
C	54	145.2	15.4	797	13	A2685845
C	55	145.2	15.4	797	13	A2685845
C	56	145.2	15.4	797	13	A2685845
C	57	145.2	15.4	797	13	A2685845
C	58	145.2	15.4	797	13	A2685845
C	59	145.2	15.4	797	13	A2685845
C	60	145.2	15.4	797	13	A2685845
C	61	145.2	15.4	797	13	A2685845
C	62	145.2	15.4	797	13	A2685845
C	63	145.2	15.4	797	13	A2685845
C	64	145.2	15.4	797	13	A2685845
C	65	145.2	15.4	797	13	A2685845
C	66	145.2	15.4	797	13	A2685845
C	67	145.2	15.4	797	13	A2685845
C	68	145.2	15.4	797	13	A2685845
C	69	145.2	15.4	797	13	A2685845
C	70	145.2	15.4	797	13	A2685845
C	71	145.2	15.4	797	13	A2685845
C	72	145.2	15.4	797	13	A2685845
C	73	145.2	15.4	797	13	A2685845
C	74	145.2	15.4	797	13	A2685845
C	75	145.2	15.4	797	13	A2685845
C	76	145.2	15.4	797	13	A2685845
C	77	145.2	15.4	797	13	A2685845

## ALIGNMENTS

```

RESULT 1
LOCUS      A0077154/c
DEFINITION
ACCESSION  A0077154
VERSION    479 bp
KEYWORDS   DNA
SOURCE     Ctr-HSP-2354D1.TF Ctr-HSP Homo sapiens genomic clone 2354D1, DNA
           sequence.
           A0077154
           A0077154.1 GI:3438338
           GSS.
           human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 479)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE      Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
JOURNAL    Unpublished (1998)
COMMENT    Other_GSSs: Ctr-HSP-2354D1.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES
            Location/Qualifiers
            source
            1..479
            /organism="Homo sapiens"

```

Email: [mouseeest@watson.wustl.edu](mailto:mouseeest@watson.wustl.edu)  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.

RESULT	3
AZ090606/c	
LOCUS	
DEFINITION	692 bp DNA
\	GSS
ACCESSION	08-MAY-2000
AZ090606	Mus musculus genomic clone RPCI-23-27B20/ DNA sequence.



VERSION	AZ090606.1	GI:7732649
KEYWORDS	GSS.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 692)	
	Zhao,S., Niernman,W., Feldblum,T., Malek,J., Shatsman,S., Akinret	
	,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P	
	and Fraser,C.M.	
TITLE	Mouse BAC End Sequences from Library Rpci-23	
JOURNAL	Unpublished (1999)	
COMMENT	Other_GSSs: Rpci-23-27&20..TJ	

Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [szhao@tigr.org](mailto:szhao@tigr.org)  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong ([pieterdejong.med.buffalo.edu](mailto:pieterdejong.med.buffalo.edu)). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([info@resgen.com](http://www.tigr.org/ldh/bac/ends/mouse/bac_end_intro.html)). BAC end page: [http://www.tigr.org/ldh/bac/ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/ldh/bac/ends/mouse/bac_end_intro.html)  
plate: 27 row: B column: 20  
Seq primer: T7  
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1..692

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPC1-23-27B20"
/clone_11p="RPC1-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
ECORI; Site_2: ECOR1; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of ECOR1 and ECOR1 Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      205 a      164 c      163 g      160 t
ORIGIN

```

Query Match	21.4%;	Score 201.8;	DB 13;	Length 692;
Best Local Similarity	58.1%;	Pred. No. 1.2e-45;		
Matches 375;	Conservative	0;	Mismatches 267;	Indels 3;
				Gaps 1

Qy	278	gctctcgagctcgtgtagccagcttccatcttcgcgccttgagggctaccgaatc	3372
Db	692	gctatgctggggtgatgacccagttcatttcctatagacacttactgogtggaaactgagtg	6533
Qy	338	ttctctcgccgcgcgcgtcccttgatagtgcttgtagctattgtgcgcctcccat	397
Db	632	tgcctccttggtgatggctttccagcgcgatgctgctgctgctgcactacacaca	5733
Qy	398	cagttatcatgaccagagactcgcgtcccaagttgtagagcgatcctctggtta	457
Db	572	ccacacattatgacaccccttctcgcacatgacattggccatctcctctgggtagggcc	5133
Qy	458	ttagtaactcaagtgtgttctctacccctgactccagctgcgcactcttgaccccat	517
Db	512	tttggaactcctcctacacagacagtcattcatgacattacgtctctgtagccatcacc	453
Qy	518	tgtatagatcaacttctctgtggaagtcctcctgacactgcgtcccaagtatacttg	577
Db	452	tga--acacattcttctgtgtagatgctgttctccctggaagctgcttgtagagacacag	386

OY	578	cagaaatggaaggcgcgaactattcccttgacgtggagcttccatactaaccoccgaaac	637
Pb	395	TGGGACAGACAACCCTACTGTGTTTGGCGTGGAAGCTGTAAATTTGGCTGTCCCTTAGACAC	336
OY	638	tcaatccatatcatcatbcttttaattgttcgcgcagacaqaltaigagaatacagctcgcgaag	697
Pb	335	TAAATTCTAAGGCACCGCATTCACACAAATTTGGCTCAATMGCAAGTGTGAAGATCAAGTCAGAATC	276
OY	698	gtcgaacaaaagcatttgggacatagtgttcccatactaattgcgtgltccttttta	757
Db	275	GCGCGAGAAAGCGTCTGGGAACTGTGGGGTCCCAOCTTACTGTGGTTTTCTTTTPTTAAG	216
OY	758	gtaacagccgtctcttgttgaacctgtcaaccacttgcgccagctccaagaaccaagaaga	817
Pb	215	GCTCACGCATGTATACACGTATCTCCAGCCTATTCACACGCTATTTGTGTAGTAAGGGAAGT	156
OY	818	tgtttctctctctcatgatgaatcattgcaccatgcataccttatccatatatacacta	877
Pb	155	TTCGCGCCCTCTTTTATATCTATCATTAATCTCAATGCTCAACCCCTGTGATTTATACCTTA	96
OY	878	ggacaagaaggagtaaagaagcgtttaaagggttggtgtcgaag	922
Pb	95	GAACACAGATGTATAAGGGGCTGTGTCAAGAGTCTGTAAAG	51
RESULT	4		
AKO17036			
LOCUS	AKO17036	1394 bp mRNA	HMC 05-JUL-2001
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:A93343B02, full insert sequence.		
ACCESSION	AKO17036		
VERSION	AKO17036.1	GI:12856091	
KEYWORDS	CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:A93343B02.		
ORGANISM	Mus musculus		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 1394)		
AUTHORS	Carninci,P. and Hayashizaki,Y.		
TITLE	High-efficiency full-length cDNA cloning methods in enzymology. 303, 19-44 (1999)		
JOURNAL	Methods in enzymology. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2 (bases 1 to 1394)		
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3 (bases 1 to 1394)		
AUTHORS	Shibata,K., Itoya,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasbiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawal,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11078861		
REFERENCE	4 (bases 1 to 1394)		
AUTHORS	The Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
REFERENCE	5 (bases 1 to 1394)		
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,		

**TITLE**  
**JOURNAL**

Arkawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Harata, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Iwama, M., Kato, H., Kawaji, T., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okeazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shitaki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Yamatsuo, M. and Hayashizaki, Y.

**Direct Submission**

Submitted (10-JUL-2000) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT**

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5' GAGAGCAGACGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot = 10.0 and subtraction to  
Rot = 100.0. Second strand cDNA was prepared with the primer  
adaptor of sequence [5'  
GAGAGCAGAGCTTCTCGATTAATTAATTAATTCCTCCCCCCCC 3']. cDNA was cleaved  
with BamHI and XhoI. cDNA of size compressed between 0.5 and 3 kb  
was selected before cloning. Vector: a modified pluscript KS(+) after  
bulk excision from lambda FLC I. Cloning sites, 5' end: SalI,  
3' end: BamHI. Host: DH10B.  
Location/Qualifiers  
1. .1394

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1894709"
/db_xref="MGI:1914036"
/cloname="4933433E02"
/sex="male"
/tissue_type="testis"
/tissue_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
332. .1282

```

```

/protein_id="BAB30564.1"
/db_xref="GI:12856092"
/translation="MEPNWSTLSEGFILVGLIDSGSEFELLCATVTLIYMLAISNGI
LLVITVDARLHPMYLLLRQLSLIDLETSVPVPTVDELNDNTISFEGCALQLE
SAMTIGVADLLLFAMAYDRVVALICHPMSPACRMLVAISMLTSLAGHST
VYTHAFPECMSCSEIRHLCEVPLTKLACAPTSOELMYVTVGFIPLPLSAITTS
SLITFVFLHMPNSGKRAIKVITCSHLTYVMEFGYAFEMVYVLESSFSPQDNITIS
FYITVTPALNLIYSLRKKEVIGAVRRKLGHIILPAATV"
BASE COUNT      355 a      337 c      304 g      358 t
ORIGIN

```

Query Match	20.9%	Score 196.6;	DB 12;	Length 1394;
Best Local Similarity	52.5%;	Pred. No. 3.9e-44;		
Matches 430;	Conservative 0;	Mismatches 389;	Indels 0;	Gaps 0;
<hr/>				
QY 103	tacactgtgacacatcttggcacaatcggacattatctagtctcagcgcgtgacaccaa	162		
Db 434	TACATGTTTGGACGTCATCAGCAATGGACGCTTACGCCGTGATCATCAGTGGATGCCGG	493		
QY 163	cttcataccccagatattttctcttaccacaatcatataactctcggatccttgtttacac	222		

Dp	494	CTTACAGTAAACCAGTACGTACCTACAGGACAGCTCTCTCATTTAGCACTCCTCTTTCACA	553
Qy	223	acatgtaacagtcgccacaacaatgctagtaaatltatgcaagcatcaagaaaglaacgatat	282
Dp	554	TCACCTTGTAACTCCCAACACACTGTTGTGATTTTCTGCTCAGAGACAAACCATATTCCTTT	613
Qy	283	cgtggcgtgtagcccaagctttccatatctttgacctgggggctactgaaatctctc	342
Dp	614	GAGGAGATGTGCCCTTTCAAATTTGTTCTCAGCAATATCATTTGGGTGTCCAGAGACCTCCTT	673
Qy	343	ctggccgtcaatgcttccttgaagttgtgtgacattctgtgcgacctccatcaaccagtt	402
Dp	674	CTGGCCTTTCATGAGGCGCTTATGATAGTAGTGGCCATTGTTCATCCTTAACTACATCATGATC	733
Qy	403	atcalgcaaccaagagactctgctccagattgycagccgcacccctgggttaactcgttttagt	462
Dp	734	TTTCATGATGCCAAAGGCGCTCAGAGCTCATGTGGCCATATCATGATGATCTTACATCTCTT	793
Qy	463	aactcagttggttgtcttaccctgaactctccagactccagctcgtcaactctgtaacctatgtata	522
Dp	794	AGTGCCCTAGGTCAACACAGTACACACATATGCACATTCCTCTTTTGGCATGTGCCAGGAAATC	853
Qy	523	gatacattctctgttgaagtcacctgcaagctcaagtcttatctgtgtgttgaagcaacaga	582
Dp	854	AGACACCTGCTCTGTGTGGATTCCTCCATGTTTAAATTTGGCTGTGTGCACACATCTCA	913
Qy	583	aatgagctgaacatcactcctctgcaatgagctctccatccataaacacctgcaactatc	642
Dp	914	TATGAGGTCATGTTTATGTGTGACAGAGATATATTTCTTATGCTGCCCTCTCTGCATTT	973
Qy	643	cttatatcatatgcttttatttctgcagagcaatgtgaagataacagctctgcaagtcga	702
Dp	974	ATTACCTCTACTACATAATTTATTTACATGCTGCTGCACATGCTTCAATATGAGGCAAG	1033
Qy	703	caaaaagcatcttggaacatgttgttccatcaactaatgtgtgtctcttttataagtata	762
Dp	1034	AAGAAAGCCCTGTGCACCTCTTCTCCCACTTACATGTGTGGATGTTCATATGGGGT	1099
Qy	763	gcgcgtctctgttgaactcgtcaacaacacttgcgcccagcttccaagaagccaagaagaagtgt	822
Dp	1094	GCCACTTTCATGATATGTGTGCGCCAGTTCCTTCCACAGTCCATAGCAACACATATTCATC	1155
Qy	823	tctctctctatggaatcaacttgtaaccatgctgaaatcccttatataatacacttgaagc	882
Dp	1154	TCTGTGTTCTATACAAATTTGCACACACAGCTCTGAACCCCTCATTTACAGCTTGAAAT	1213
Qy	883	aagagagtaagaagaagctttaaagttgtgtgtcaga	921
Dp	1214	AAGGAGGTGATTTGGAGCTGTTTAAAGAGTATCTGGGAGA	1252

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PubMed	REFERENCE	AUTHORS
AK016560	3063 bp mRNA	AK016560	1	GI:12853537	mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:4932441H21, full insert sequence.	Mus musculus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus	Carninci, P. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning	Methods in enzymology. 303, 19-44 (1999)	95278233	10349636	2 (pages 1 to 3063)	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,



Db 1260 TCTGTGTTCTATACATTTGTCACACCGCTGACACCCCTCATTTACAGCTGAGAAAT 1319  
 QY 883 aagagagtaagaagagccttaagaaggttggtgcaga 921  
 Db 1320 AAGGAGGTGATTGAGCTGTTAGAGAGTACTGGGAGA 1358

## RESULT 6

A2969227

LOCUS A2969227 642 bp DNA GSS 27-APR-2001  
 DEFINITION 2M024J124R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
 clone UUGC2M024J124 R, DNA sequence.

A2969227

ACCESSION A2969227  
 VERSION A2969227.1 GI:13840454  
 KEYWORDS GSS

SOURCE

ORGANISM house mouse.  
 Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 642)

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weis, R.

Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

TITLE

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss  
 University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0241 row: J column: 24

Seq primer: CACACGAGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 642.

Location/Qualifiers

FEATURES

SOURCE

1. 642  
 /organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M024J124"

/clone\_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (female) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PMD42 (g11473211419b|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT  
 ORIGIN

130 a 166 c 129 g 217 t

Query Match 20.8%; Score 195.6; DB 13; Length 642;  
 Best Local Similarity 56.5%; Pred. No. 6.2e-44;  
 Matches 363; Conservative 0; Mismatches 279; Indels 0; Gaps 0;

QY 233 tcccaacaatgctagtaaatatgcaagcaltcaagaagtaacgtatcgtgctg 292  
 Db 1 tccccccagatgagtgatcttctgtcagagagaaatcatttccgtgagagctg 60  
 QY 293 tagccagcttctcaatattctgacctgagggtctactgaatattctccgacctga 352  
 Db 61 tgaaccagctgttgatcatttctgtggtgctcagagagtgctccctggcagca 120  
 QY 353 tgccttatagagtggtgagctatttgcggcctctccattaccagatcacaagcacc 412  
 Db 121 tggcatatgcatgatatattgctatctgtaaccggttaagagatcattattatgaca 180  
 QY 413 agagactcgcctcagttgagcagccagcactcctggtggttactggtttagaactcag 472  
 Db 181 agcgtctgagcagctgagtgagcagcttcatgctggagcagtggttctcactgactg 240  
 QY 473 ggtgtctcctcctgactctccagctgcagctctggtgagccctatgtatgatacttc 532  
 Db 241 tgcacacggttttgaccttgcacctgccccttgggttaacatgacatcattatttct 300  
 QY 533 tctgtaagctccctgcactgcaagctatctctggtgagcaacagcaagtgaggtcgt 592  
 Db 301 tctgtgacatgactcctcctgctgctgctgctgctgctgctgctgctgctgctg 360  
 QY 593 aactatcctgctcagtgagcctctccatcctatcctccctgcaactcattatcct 652  
 Db 361 ctttgctgctgcaatgagatcctgagctgagctgctgctgctgctgctgctgctg 420  
 QY 653 agcttttatgctcagagagatgagatcagctcgtcgtgaagtgagcaaaagct 712  
 Db 421 acccttaccatgctgctcagcagctgagagatcgttctgctgagagagcagcaagc 480  
 QY 713 ttggagacatggtgtcccaatcctaatgtgtctcttttatagtaagcagctctcgt 772  
 Db 481 ttccacagctgctcctccagcagctgctgctgctgctgctgctgctgctgctgct 540  
 QY 773 tttacctgcaacacctcgcagcagctcacaagacaaagatggttctctctct 832  
 Db 541 cgtatgtgagggccacatcttcttcttcttcttcttcttcttcttcttcttct 600  
 QY 833 atggaatcatgctaccatgctgaatccctctatatacac 874  
 Db 601 atagctgtgtcagacacctgagatctgtaatttattatgcg 642

## RESULT 7

A2511623/c

LOCUS A2511623 632 bp DNA GSS 05-OCT-2000  
 DEFINITION 1M0356C17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0356C17 R, DNA sequence.

A2511623

ACCESSION A2511623  
 VERSION A2511623.1 GI:10692939  
 KEYWORDS GSS

SOURCE

ORGANISM

house mouse.  
 Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 632)

REFERENCE

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
 and Wright, D., Weis, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

COMMENT

Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT



QY 875 ttgagcaaggaaggttaagaagctttaaaggttgctgcgaagctcttctaataca 934  
 |||  
 DB 208 tttagcaaggaaggttaagaagctttaaaggttgctgcgaagctcttctaataca 149  
 |||  
 QY 935 agaataa 942  
 |||  
 DB 148 ACAATAA 141

## RESULT 9

A2694245

LOCUS 326 bp DNA GSS 18-DEC-2000  
 DEFINITION AS1-2HBG3005 Genetrap HL-60 Human Promyelocytic Leukemia Library  
 Homo sapiens genomic 5', DNA sequence.

ACCESSION

A2694245

VERSION

A2694245.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 326)

AUTHORS

Henkel, G., Llyanage, M., Pratt, E., Huang, D., Riley, M., Bernardino, A., Durick, K. and Pollok, B.

TITLE

Exon-trap tags from a HL-60 Genomescreen(TM) Library

JOURNAL

Unpublished (2000)

COMMENT

Contact: Greg Henkel

Gene Expression

Autora Biosciences Corp.

11010 Torreyana Road, San Diego, CA 92121, USA

Tel: 8584048436

Fax: 8584046719

Email: henkel@autora.bio.com

Pools of cells were isolated from a Genomescreen(TM) library. The

library of cells was generated by retroviral integration of a gene

tagging element consisting of: 1) A promoterless beta-lactamase

proceeded by a splice acceptor as a reporter for gene expression;

2) A promoter driving neomycin resistance followed by a splice

donor to trap downstream exons. 3' RACE from neomycin gene was

performed using total RNA from isolated pools. Output was shotgun

cloned in pAMP-1 and used to transform DH5-alpha competent

bacteria. 5' ends of reported sequences were immediately preceded

by splice donor from the trapping construct.

Class: exon-trapped.

Location/Qualifiers

1..326

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Genetrap HL-60 Human Promyelocytic Leukemia

Library"

/lissue\_type="acute promyelocytic leukemia"

/cell\_type="promyeloblast"

/cell\_line="HL-60"

/note="Organ: peripheral blood; Vector: pAMP-1; 3' RACE of

total RNA from genetrap pools; shotgun clone in pAMP-1 and

used to transform DH5-alpha competent bacteria."

BASE COUNT

109 a 65 c 57 g 95 t

ORIGIN

Query Match

Best Local Similarity

100.0%; Score 184; DB 13; Length 326;

Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 tacagccgtctctgtacacgaacacgttcgcccagctccaaaggaagaagaat 818

DB 1 TACAGCCGTCTCTGTACACGAACACGTTCCGCCCAAGGCAAGGAAGAT 60

QY 819 ggttctctctctatgaatcatgtgacccatgctgaatcccttatataacttag 878

DB 61 GGTTCCTCTCTTATGGAATCATGTGACCATGCTGATATATATACACTTAG 120

QY 879 gaacaaggaaggttaagaagctttaaaggttgctgcgaagctcttctaataca 938

DB 121 GAACAAGGAGGTAAAGGAGGCTTTAAAGGTTGCTGCAAGAGCTTCTTATACAGAA 180  
 |||  
 QY 939 ataa 942  
 |||  
 DB 181 ATAA 184

## RESULT 10

BH111304

LOCUS 788 bp DNA GSS 19-JUL-2001  
 DEFINITION RPI-24-367N6.TJ RPI-24 Mus musculus genomic clone RPI-24-367N6,  
 DNA sequence.

ACCESSION

BH111304

VERSION

BH111304.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 788)

AUTHORS

Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akimov, B., Levins, M., Tseng, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.

TITLE

Mouse BAC End Sequences from Library RPI-24

JOURNAL

Unpublished (1999)

COMMENT

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@tigr.org). Clones may be purchased from BACPAC

Resources (<http://www.chori.org/bacpac/orderingframe.html>). BAC endpage: [http://www.tigr.org/tigr/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tigr/bac_ends/mouse/bac_end_intro.html)

Plate: 367 row: N column: 6

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..788

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPI-24-367N6"

/clone\_lib="RPI-24"

/sex="male"

/cell\_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI. The

RPI-24 Mouse BAC library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MhoI partially digested male C57BL/6J

DNA."

BASE COUNT

168 a 217 c 162 g 241 t

ORIGIN

Query Match

Best Local Similarity

19.2%; Score 180.4; DB 13; Length 788;

Matches 413; Conservative 0; Mismatches 351; Indels 2; Gaps 2;

QY 103 tacagtgtacacgttcgtggaacatcgtccattatctcaagtgacagctgtgacacaa 162

DB 6 TACAGTGTACACGTCTGCAAAATGAGATTAATCTTAGTATGATCTGCTGACCCCAAG 65

QY 163 ctcataccacatgtatcttcttcaacatcatcaaccctggatctgttaacc 222

DB 66 CTGACACCCCTATGATCTTCTTCTTCTGATCTGCGCATTTCTTGACATATCTTATGCT 125

QY 223 acatgtacagtcacacaaatctagtaattatcatcagcatcaggaagtaacagtat 282

This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1. .640  
/organism="Homo sapiens"  
/db\_xref="GDB:7634182"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-350623"  
/clone\_id="RPC1-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Rpci11 Human Male BAC library"

BASE COUNT 185 a 99 c 140 g 216 t

ORIGIN

Query Match 19.0%; Score 178.6; DB 13; Length 640;  
Best Local Similarity 93.0%; Pred. No. 3.5e-39;  
Matches 198; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Db 640 ATCTAATGTTGGTGCACATTTTATGTGTACAGCATTCATGTACCTGACACCACTT 581  
|||||  
QY 731 atcaaatgtgtgtctcttttatagtaagccgtctcgtgtactgaaccactt 790  
|||||  
Db 640 ATCTAATGTTGGTGCACATTTTATGTGTACAGCATTCATGTACCTGACACCACTT 581  
|||||  
QY 791 ccccaagctcccaagcaagaaagatggtttctctctctatgaaatcatgtcacca 850  
|||||  
Db 580 CACCAGCTCCCAAGACCGGGGAAGAAGTGTTCTCTCTCTGTGGAATCATGTGACCCA 521  
|||||  
QY 851 tgcgtaatcccttatataatacacttagaacaagaggtataagaaagccttaaaagt 910  
|||||  
Db 520 TGCGAATCCCTTATATATACACTTAGACAAAGAGTAAGAGCACTTTAAAGT 461  
|||||  
QY 911 tggttgc-aagagcttcttctaataagaataa 942  
|||||  
Db 460 TGGTTGCAAAAGTCTTCTTAATCAAGAAATTA 428  
|||||

RESULT 12  
LOCUS A2413817 588 bp DNA GSS 03-OCT-2000  
DEFINITION M0188F05F mouse 10kb plasmid UGCG1M library Mus musculus genomic  
clone UGCG1M0188F05 F, DNA sequence.  
ACCESSION A2413817  
VERSION A2413817.1 GI:10537830  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 588)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0188 row: F column: 05  
Seq primer: CGTGTAAACGACGCCACAGT  
Class: plasmid ends  
High quality sequence stop: 588.

This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1. .640  
/organism="Homo sapiens"  
/db\_xref="GDB:7634182"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-350623"  
/clone\_id="RPC1-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Rpci11 Human Male BAC library"

BASE COUNT 185 a 99 c 140 g 216 t

ORIGIN

Query Match 19.0%; Score 178.6; DB 13; Length 640;  
Best Local Similarity 93.0%; Pred. No. 3.5e-39;  
Matches 198; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Db 640 ATCTAATGTTGGTGCACATTTTATGTGTACAGCATCTCATGTACCTGACACCATTT 581  
|||||  
QY 731 atcaaatgtgtgtctcttttatagtaagccgtctcgtgtactgtgaaccactt 790  
|||||  
Db 640 ATCTAATGTTGGTGCACATTTTATGTGTACAGCATCTCATGTACCTGACACCATTT 581  
|||||  
QY 791 ccccaagctcccaagcaagcaagaagatggtttctctctctatagaaatcattgaccca 850  
|||||  
Db 580 CACCAGCTCCCAAGACCGGGGAAAGATGGTTCTCTCTCTGTGGAATCATGTGACCCA 521  
|||||  
QY 851 tgcgtaatcccttatataatacacttagaacaagaggtataagaaagccttaaaagt 910  
|||||  
Db 520 TGCGAATCCCTTATATATACACTTATAGAACAAAGAGTAAGAGACCTTTAAAGT 461  
|||||  
QY 911 tggttgc-aagagcttcttctaataagaataa 942  
|||||  
Db 460 TGGTTGCAAAAGTCTTCTTAATCAAGAAATTA 428  
|||||

RESULT 12  
LOCUS A2413817 588 bp DNA GSS 03-OCT-2000  
DEFINITION M0188F05F mouse 10kb plasmid UGCG1M library Mus musculus genomic  
clone UGCG1M0188F05 F, DNA sequence.  
ACCESSION A2413817  
VERSION A2413817.1 GI:10537830  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 588)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0188 row: F column: 05  
Seq primer: CGTGTAAACGACGCCACAGT  
Class: plasmid ends  
High quality sequence stop: 588.



## FEATURES

source

Location/Qualifiers

1. 588

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUCG1M018F05"

/clone\_lib="Mouse 10kb plasmid UUCG1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 157 a 139 c 104 g 188 t

## BASE COUNT

157 a 139 c 104 g 188 t

## Query Match

18.6% Score 175.6; DB 13; Length 588;

Best Local Similarity 66.2% Pred. No. 2.4e-38;

Matches 253; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

31 gagttatctcgtggttctcagatcgacctgtgagcttccactccttctgtc 90

198 GACTTCATCCTCTCGGATTTCCAGTCGACCCCAACGTGACACATCATCTCGAGTT 257

91 tctctgattcttaccagctgtgacatccttggcaatcgtgacattatctagtcacgc 150

258 GCTTTGCTTCTATATATGTGACTGTGGAGAACACAAACATATCTACTGTCCTAT 317

151 ctggacaccaactctacacatgtaatttcttaccatcactcctcgtgact 210

318 CTGGACTCTCAGCTCCATATCTCCATGTAATTTCTTATTAATTTGCTTTCGTGAC 377

211 ctgtgtacaccacatgtacagtcacacaaatgtagtaattttagcagcatcaggaa 270

378 CTCTGTATTAATTAATGATGTCACAGATGCTGTAATATCTATGGGCCCAAGAAG 437

271 gtaacagattacgtgctgtgtagccagcttccatattctcgtgctgggagctact 330

438 TCTATTAATTAATGAGGAGGCTGTACTCAAGTCTTCTTCCCTGGACCTGGAGCCACA 497

331 gaatatcttcctcgtgacgtatcgtttagtagttagttagttagttagttagttagt 390

498 GAATGTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 557

391 cactacagttatcatgacac 412

558 CACTACAGTAAATTAATGACAC 579

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

## SOURCE

ORGANISM

house mouse.  
Mus musculus

REFERENCE

AUTHORS

1 (bases 1 to 740)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly

M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

JOURNAL

COMMENT

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah

Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0429 row: N column: 15

Seq primer: CACACGAGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 740.

Location/Qualifiers

1. 740

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUCG1M0429N15"

/clone\_lib="Mouse 10kb plasmid UUCG1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b1AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 276 a 134 c 159 g 171 t

## BASE COUNT

276 a 134 c 159 g 171 t

## Query Match

18.3% Score 172; DB 13; Length 740;

Best Local Similarity 53.9% Pred. No. 2.5e-37;

Matches 375; Conservative 0; Mismatches 320; Indels 1; Gaps 1;

124 aatctgacattatctcgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgt 183

697 AATCTTGAATGATCAATTTTGAATGATCAATTTTGAATGATCAATTTTGAATGATCAAT 638

184 ttcttaccatctacatcactcctggtatcttctacacacatgtagtccacaaatg 243

637 TTTCTCAGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 578

244 ctatgaaattatgacagcatcaggaagtaacatgtagtctggtctgtagccagctt 303

577 CTTGTAGACACTGCTGTAAGAAATACATTAATCTCTTTCTTGGCTGTGCTATGCAAGTTC 518

Db

Qy

Db

Qy

Db

Qy

Db



```

OY 304 ttcaatattctgcttgaggagctactgaataatcttcctccgctcaatgctcttgaat 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 517 TTCACCTTCTGATTTTATATAGATGCTGAGTGTGCTGCTGCGAGTATGTCATTTGAT 458
OY 364 aggtttgaagattatgttcgagctctccatctactcaatgatalcatgaccagagactcgc 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 457 CGATATTAAGGCATTAAGTAAACCCCTTTGATGAGATGAGACATGTCAGGAAGTGTGC 398
OY 424 ctccagcttgagcagccgcaatccctgggttaactggttttaactcaatgctgctgctacc 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 TTCACATTAATGACCTGGTGTATTTATCTGTGGCATTAAGAGATGCTTTGATACATACAA 338
OY 484 ctgactcccaagctgcacactgtgacccctatgtagatagactctctctgtagaagc 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 337 TTGACATTCACACTTATGTTTCTGTGGGCGAATGAGATTAATCATTTCTTCTGTGATANT 278
OY 544 cctgacactgctcaatgatalctatgtgttgtagaacaacagcaaatgaggtcgaactatcc 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 277 CCTCCAGTGTAGTACTGCTGTGTGACAGACAGGTCATGATTTTACTCATATTCACT 218
OY 604 gtcaatgagctcttccatctaatlaccctcgacatccatctctatataatgcttatt 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 217 GTTTTGTGTTTCAATGAACTGAGCACACATCTGAGAGTTCTTATCTCTACTGTTACATC 158
OY 664 gtccagagagatltgagagatacagctcgtgagagtcgacaaagacattgagacatgt 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 157 ATCTCATGATGTTCTGAAGATGATGTTCTGTGCTGTGGAGGTTAAAGCTTCTCTACCTGT 98
OY 724 gttcccaatcaattgtgtgtcctcttttataagtaagccgctctgtgtgactgcaa 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97 ACATCACACACCTGACGCTGTTGCAATTTTTCAGGAACATGATCTCTTCAATGA-TTCCCG 39
OY 784 ccaacttcgcccagctccaagcaaggaagaatgt 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38 CCAAGTCTCTCTACTCCTAGATCAAGATTAAGTG 3

```

```

RESULT 14
LOCUS A2271125 548 bp DNA GSS 26-JUL-2000
DEFINITION RPCI-23-146J4.TV RPCI-23 Mus musculus genomic clone RPCI-23-146J4,
VERSION A2271125
KEYWORDS A2271125.1 GI:9484742
SOURCE GSS.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 548)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akimov,
B., Levins,M., McGinn,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
unpublished (1999)
Other GSSs: RPCI-23-146J4.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pleter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderinframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 146 row: J column: 4
Seq primer: T7
Class: BAC ends.
Location/Qualifiers

```

FEATURES

```

source
1..548
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-146J4"
/clone_11b="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site:1:
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies).
BASE COUNT 147 a 124 c 134 g 143 t
ORIGIN

```

```

Query Match 17.4% Score 164; DB 13; Length 548;
Best Local Similarity 61.9%; Pred. No. 4.1e-35;
Matches 260; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

```

```

OY 517 gtgataagatcacttctctgtgaagtcctgcaactgctcaagtatctgtgtgagaca 576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 470 GTAATAGACCATTTCTCTGTGAGAGCCAGCCCTTTGAACATGACTGCACTGACACA 411
OY 577 acagcaaatgagctgcaactatctcctgtcagtgagctctccatctaataccctgaca 636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 410 AGTGTGAATGAAGCTGAGCTTAATGTTTGGAGCTTGTCTCTCTGCTGCTCAAC 351
OY 637 ctcacacctatatacatatgcttctgttcgagagagatltgagataagctcgtcga 696
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 350 CTGATCTGCGGTAACCTATGTTGTCATGCTCAAGCAAGTCAAACTCCGTTGCTGAG 291
OY 697 ggtcgacaaaagaagcatlttgagacatgtgtgtccatctaatgtgtgtcttatt 756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 290 AGTCGCGGGAAGCATTTAATACGTCGCTTACATCTGCTGTGCTGCCCTTCTTCTAT 231
OY 757 agtaagccgctctgtgtactgcaaccacttgccagctccagatccaaggaagaag 816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 230 TTCACAGCATATGATGATGATGTCACCTCCCTCAAGCTCATCTCAAGAAAAAAG 171
OY 817 atggttctctctctatgtaaatcttgacccaatgctgaatcccttatataact 876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 170 ATCATGGCTCTGTCTATGCAATGTGCACACCTACCTCAACCATTCATCTACCTTG 111
OY 877 aggaacaagagagtaagaagagctttaaaggttggtgcaagagctctcttaacag 936
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 110 AGGAATTAAGCATGTTAAGGCTCCCTGAGAGAGGCACTAACAAAGAGTTTGGGTCAAG 51

```

```

RESULT 15
LOCUS A2765752 632 bp DNA GSS 16-FEB-2001
DEFINITION IM0562E19R Mouse 10kb plasmid UNGC1M library Mus musculus genomic
VERSION A2765752
KEYWORDS A2765752.1 GI:12882100
SOURCE GSS.
ORGANISM house mouse.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 632)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenan,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid Inserts
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL Contact: Robert B. Weiss
COMMENT unpublished (2000)

```



Tue Feb 26 09:22:08 2002

us-09-755-017-1.rst



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 03:41:12 ; Search time 1356.19 seconds

(without alignments)  
11458.824 Million cell updates/sec

Title: US-09-755-017-1

Perfect score: 942  
Sequence: 1 atgaattgggttaatgacag.....tcttcttaataagaataaa 942

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_hlg: \*  
3: gb\_in: \*  
4: gb\_ov: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rnd: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rnd: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	940.4	99.8	44788	2	AL133267	AL133267 Human DNA
2	940.4	99.8	176277	2	AC025941	AC025941 Homo sapi
3	727.8	77.3	1074	9	HS302593	AJ302593 Homo sapi
4	726.2	77.1	1074	9	HS302584	AJ302584 Homo sapi
5	726.2	77.1	1074	9	HS302585	AJ302585 Homo sapi
6	726.2	77.1	1074	9	HS302586	AJ302586 Homo sapi
7	726.2	77.1	1074	9	HS302587	AJ302587 Homo sapi
8	726.2	77.1	1074	9	HS302588	AJ302588 Homo sapi
9	726.2	77.1	1074	9	HS302589	AJ302589 Homo sapi
10	726.2	77.1	1074	9	HS302590	AJ302590 Homo sapi
11	726.2	77.1	1074	9	HS302591	AJ302591 Homo sapi
12	726.2	77.1	1074	9	HS302592	AJ302592 Homo sapi
13	726.2	77.1	100375	9	HS193B12	298744 Human DNA S
14	726.2	77.1	166758	2	AC024428	AC024428 Homo sapi
15	726.2	77.1	176277	2	AC025941	AC025941 Homo sapi
16	665.2	70.6	185523	2	AL589742	AL589742 Mus muscu
17	665.2	70.6	222824	2	AL589651	AL589651 Mus muscu
18	647.4	68.7	649	9	U86270	U86270 Homo sapien
19	630.6	66.9	649	9	U86271	U86271 Homo sapien
20	607	64.4	646	9	U86275	U86275 Homo sapien
21	599.6	63.7	222824	2	AL589651	AL589651 Mus muscu
22	588.4	62.5	1214	10	RATOLIRCE	L34074 Rat OLI rec
23	538.8	57.2	942	9	HS302546	AJ302546 Homo sapi
24	537.2	57.0	942	9	HS302537	AJ302537 Homo sapi
25	537.2	57.0	942	9	HS302538	AJ302538 Homo sapi
26	537.2	57.0	942	9	HS302539	AJ302539 Homo sapi
27	537.2	57.0	942	9	HS302540	AJ302540 Homo sapi
28	537.2	57.0	942	9	HS302541	AJ302541 Homo sapi
29	537.2	57.0	942	9	HS302542	AJ302542 Homo sapi
30	537.2	57.0	942	9	HS302543	AJ302543 Homo sapi
31	537.2	57.0	942	9	HS302544	AJ302544 Homo sapi
32	537.2	57.0	942	9	HS302545	AJ302545 Homo sapi
33	537.2	57.0	942	9	HS80119	AL022727 Human DNA
34	397.2	42.0	144868	2	AC011571	AC011571 Homo sapi
35	393.2	41.7	130279	2	HS302541	AL121944 Human DNA
36	380	40.3	185302	2	AL359352	AL359352 Mus muscu
37	380	40.3	251546	2	AL365336	AL365336 Mus muscu
38	370	39.3	178295	33	AL390860	AL390860 Human DNA
39	370	39.3	180657	2	AC091612	AC091612 Homo sapi
40	368.4	39.1	190889	2	AL357039	AL357039 Homo sapi
41	365.8	38.8	939	9	HS302565	AJ302565 Homo sapi
42	365.8	38.8	939	9	HS302566	AJ302566 Homo sapi
43	364.2	38.7	939	9	HS302555	AJ302555 Homo sapi
44	364.2	38.7	939	9	HS302560	AJ302560 Homo sapi
45	364.2	38.7	939	9	HS302561	AJ302561 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS AL133267  
DEFINITION Human DNA sequence from clone RP3-408B20 on chromosome 6 Contains ESTs, STSs and GSSs. Contains a gene and two pseudogenes for novel 7 transmembrane receptors (Olfactory family) and a gene for a novel protein similar to 60S acidic ribosomal protein p2 (RPLp2), complete sequence.

ACCESSION AL133267  
VERSION AL133267.9 GI:10185396  
KEYWORDS HTG: 7 transmembrane; Olfactory receptor; ribosomal protein; RPLP2.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 44788)  
AUTHORS Williams, S.  
TITLE Direct Submision  
JOURNAL Submitted (25-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

## COMMENT

requests: clonerequest@sanger.ac.uk  
On Sep 18, 2000 this sequence version replaced gl:9588470.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
<http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr6>  
RP3-408B20 is from the library RPI-3 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see <http://dacpac.med.buffalo.edu/>  
VECTOR: pcypac2  
IMPORTANT: This sequence is not the entire insert of clone  
RP3-408B20. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone RPI-31316 is at 44689 in this sequence.  
The true right end of clone RPI-193B12 is at 100 in this sequence.  
Location/Qualifiers

## FEATURES

## source

1. 44788

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="6"

/clone="RP3-408B20"

/clone.lib="RPI-3"

complement(1..100)

## misc\_feature

/note="match: STS: Em:HS193B12"

## repeat\_region

222..535

## gene

/note="AluX repeat: matches 3..301 of consensus"

## CDS

932..1870

/gene="AJ408B20.1"

/note="AJ408B20.1 (novel 7 transmembrane receptor  
(olfactory family) (hs6M1-33p) pseudogene)"match: CDNAS: Em:M64392 Em:AF102516 Em:AF044033 Em:AF102525  
Em:X89699 Em:AF102521 Em:AF042016 Em:AF044033 Em:AF102525  
Em:AF044034 Em:AF044035 Em:AJ003145 Em:Y14442 Em:AF102537  
Em:Y10529 Em:AF102529 Em:AF044038 Em:AF015165 Em:Y07557  
Em:X89706 Em:AF042020 Em:M64391 Em:AF102532 Em:AF102533  
Em:M64376 Em:AF102535 Em:M64377 Em:M64378 Em:AF102537  
Em:AF091573 Em:AF091574 Em:AF091577 Em:AF091578  
Em:AF102540 Em:M64385 Em:AF044053 Em:M64386 Em:M64387  
Em:M64388 Em:AF091580 Em:DJ12820match: proteins: Tr:076000 Tr:076001 Tr:076002 Tr:035434  
Sw:015062 Tr:09WVD9 Sw:013607 Tr:090806 Tr:063394  
Tr:09WV13 Tr:09WV19 Tr:09T084 Tr:095006 Tr:09Y393  
Tr:09Q222 Sw:095156 Sw:095157 Tr:09UGF6 Tr:09Y299  
Tr:09UJ15 Tr:062007 Sw:P33275 Tr:095371 Tr:09UJ40  
Tr:09Q218 Tr:09Q219 Tr:09PSJ3 Tr:070526 Tr:09Q221"

/codon\_start=1

/pseudo

/evidence="not\_experimental  
complement(1518..1963)"

## misc\_feature

/note="match: GSS: Em:AQ456128"

## repeat\_region

2024..2298

## repeat\_region

/note="Alu repeat: matches 1..280 of consensus"

## repeat\_region

2299..2382

## repeat\_region

/note="42 copies 2 mer ag 79% conserved"

## repeat\_region

2559..2650

/note="2 copies 46 mer 91% conserved"

## misc\_feature

complement(4227..4576)

## repeat\_region

/note="match: STS: Em:G31214"

## misc\_feature

/note="L2 repeat: matches 2375..2750 of consensus"

## repeat\_region

complement(4834..5686)

## repeat\_region

/note="match: GSS: Em:AQ748194"

## repeat\_region

5098..5236

## repeat\_region

/note="L2 repeat: matches 1381..1525 of consensus"

## repeat\_region

5707..5843

## repeat\_region

/note="MIR repeat: matches 102..240 of consensus"

## repeat\_region

6275..6509

## repeat\_region

/note="L1M4 repeat: matches 5093..5329 of consensus"

## repeat\_region

complement(6590..6889)

## repeat\_region

/note="match: GSS: Em:U94441"

## repeat\_region

6679..6752

## repeat\_region

/note="MER76 repeat: matches 613..685 of consensus"

## repeat\_region

6753..6850

## repeat\_region

/note="L1P repeat: matches 1..98 of consensus"

## repeat\_region

6820..7010

## repeat\_region

/note="L1P repeat: matches 580..776 of consensus"

## repeat\_region

7006..7615

## repeat\_region

/note="L1P repeat: matches 900..1509 of consensus"

## repeat\_region

7611..9145

## repeat\_region

/note="L1P repeat: matches 3813..5344 of consensus"

## repeat\_region

9751..10411

## repeat\_region

/note="HERV repeat: matches 3244..3919 of consensus"

## repeat\_region

10535..10663

## repeat\_region

/note="L1R16A repeat: matches 203..307 of consensus"

## repeat\_region

10964..11131

## repeat\_region

/note="MLT1H repeat: matches 1..168 of consensus"

## repeat\_region

11705..12141

## repeat\_region

/note="L1M4 repeat: matches 3898..4367 of consensus"

## repeat\_region

13143..13516

## repeat\_region

/note="match: GSS: Em:AQ100659"

## repeat\_region

13668..13803

## repeat\_region

/note="MIR repeat: matches 58..199 of consensus"

## repeat\_region

complement(15225..15506)

## repeat\_region

/note="match: GSS: Em:AQ100750"

## repeat\_region

15627..16228

## repeat\_region

/note="match: GSS: Em:AQ544878"

## repeat\_region

16824..17089

## repeat\_region

/note="Alu repeat: matches 29..311 of consensus"

## repeat\_region

17090..17158

## repeat\_region

/note="Alu/PluM repeat: matches 1..81 of consensus"

## repeat\_region

17533..17848

## repeat\_region

/note="AluX repeat: matches 3..298 of consensus"

## repeat\_region

17789..18240

## repeat\_region

/note="match: STS: Em:G31213"

## repeat\_region

20021..20249

## repeat\_region

/note="L1P repeat: matches 5904..6155 of consensus"

## repeat\_region

20708..21649

## repeat\_region

/gene="AJ408B20.2"

## repeat\_region

20708..21649

## repeat\_region

/note="match: CDNAS: Em:AF044034 Em:Y14442 Em:X89700  
Em:Y07557 Em:AF042023 Em:AF091573 Em:AF091574 Em:X89670  
Em:M64391 Em:AF102540 Em:AF102533 Em:M64388 Em:DJ12820  
Em:AF034896

## repeat\_region

match: ESTs: Em:M68399

## repeat\_region

Tr:063394 Tr:095499 Tr:09Q219 Sw:P23270 Tr:09Y397  
Tr:09Q222 Tr:09Y299 Tr:062943 Tr:090807 Tr:090808  
Tr:09WV13 Tr:063395 Tr:095918 Sw:P23266 Sw:P37069  
Sw:095156 Tr:09UGF4 Tr:09WU86 Tr:060403 Tr:076000  
Sw:P37070 Tr:09QW38 Tr:077758 Tr:09Q217 Tr:09R0K3  
Tr:09R0K4 Tr:060412 Tr:09Q700 Tr:09QW36 Tr:09UJ40  
Sw:P37072 Sw:P37071 Tr:09Y395"

## repeat\_region

/codon\_start=1

## repeat\_region

/evidence="not\_experimental  
/product="AJ408B20.2 (novel 7 transmembrane receptor  
(olfactory family) (hs6M1-32))"

```

/protein_id="CAC14518.1"
/db_xref="gi:10944516"
/translation="MNNVNDISIODEFILGSDRPMLEFLLVFLVLSYTTITGNLT
ILVSRIDTRKLTHTPMFFLNLSDLCYTCVPMVLNVCSTRKVISYGVQDLF
IFLALGTEYLILLAVMSDFRVAICRPLHSVIMHORICLOLAASVNTGSNWSL
TLTQLPLCDPYVDHFLCEAFPLKLSCTETANEAEFLVSLFLPLTLLISY
AFIVRAVIRIOASGRKAFGTGSHLIVSFLSTAVSYLQPPSPSKDQKMWSL
FYGIAPMLNDPLIYTLANKKEGFRKLVARVFLIKK"
repeat_region
/notes="LtrR15 repeat: matches 127 . 212 of consensus"
/notes="complement(21244 . 21671)"
misc_feature
/notes="match: GSS: Em:A0077154"
21964 . 21999
repeat_region
/notes="18 copies 2 mer ca 100% conserved"
23289 . 23944
repeat_region
/notes="LIMB2 repeat: matches 5508 . 6168 of consensus"
24195 . 24291
repeat_region
/notes="FLAM_C repeat: matches 38 . 136 of consensus"
24448 . 24781
misc_feature
/notes="match: GSS: Em:B75689"
24537 . 24857
misc_feature
/notes="match: GSS: Em:A0102724"
24739 . 25051
repeat_region
/notes="Alusq repeat: matches 1 . 310 of consensus"
25407 . 25944
repeat_region
/notes="LIM4 repeat: matches 5206 . 5763 of consensus"
25945 . 26050
repeat_region
/notes="Alusg/x repeat: matches 195 . 302 of consensus"
26055 . 26632
misc_feature
/notes="match: GSS: Em:A0057601"
26069 . 26126
repeat_region
/notes="LIMB7 repeat: matches 6111 . 6169 of consensus"
26164 . 26220
repeat_region
/notes="LIMB7 repeat: matches 6111 . 6168 of consensus"
26221 . 26391
repeat_region
/notes="Alubd repeat: matches 137 . 306 of consensus"

Query Match          99.8%  Score 940.4: DB 9: Length 44788;
Best Local Similarity 99.9%  Pred. No. 1.4e-258;
Matches 941: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

481 accctgactctcagctgcgaactctgtgaccctatgtgatagatcactctctgtaa 540
|||||
21168 ACCCTGACTCTTCCAGCTGCACTGTGTGACCCCTATGTGATGATCACTTTCTGTGAA 21247
|||||
541 gtccctgactgactgaattatctgtgtgtgagacaagaacaatgagctgaatttc 600
|||||
21248 GTCCCTGCACTGCTCAAGTATCTGTGTGAGACAACAGCAATGAGCTCACTATTC 21307
|||||
601 ctgtcagtgagctcttcatactaatbaacctgacactcatcctatatcatatgtctt 660
|||||
21308 CTGTGCACTGAGCTCTTCCATCTAATTAACCCCGACACATCATCTATATCATATCTTTT 21367
|||||
661 attgtccgagagatitagaagatacagctgtgtgagagctgacaaagacatttgagca 720
|||||
21368 ATTGTCCGAGAGATTTAGAGATACAGTCTGTGAGAGTCCACAAAGCATTTGGGACA 21427
|||||
721 tgtgtctccatctaatgtgtgtgtctctttttataglacagccgtctctgtactgt 780
|||||
21428 TGTGTTCCCATCTAATGTGTGTCTCTTTTATAGACAGCGCTCTGTGTACCG 21487
|||||
781 caaccacacttgcagctccagctccaaagacaaagaaagatggtctctctctcatgatac 840
|||||
21488 CAACCACCTTCCGCACTCCCAAGACCAAGAAAGATGTTCTCTCTATGTGAATC 21547
|||||
841 attgcaccatgctgaatcccttatatacacttagaacaagagtgaaagagagc 900
|||||
21548 ATTGCACCATCTGCTAATCTCCTTATATATACACTTAGAACAAGAGGTAAGAGAGGC 21607
|||||
901 tttaaaagtggtgtgcaagagctctcttaacaaagaataa 942
|||||
21608 TTAAAGGTGTGTGCAAGAGCTCTTATATCAAGAATAA 21649
|||||

RESULT 2
AC025941 176277 bp DNA HTG 26-MAY-2000
LOCUS Homo sapiens chromosome 6 clone RP11-635011 map 6, WORKING DRAFT
DEFINITION AC025941.2 GI:7408057
VERSION AC025941.2 GI:7408057
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 176277)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 6, clone RP11-635011
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176277)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
Boguslavsky,L., Bouhassira,D., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Chovel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., DeArnell,R., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardy,J., Glend,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lechoczy,J.,
Levine,R., Lien,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McNeeters,R.,
Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testa,E., Theodore,J., Threlk,A., Travers,M., Trifillio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
TITLE
JOURNAL
Submitted (17-MAR-2000) Whitehead Institute/MIT Center for Genome

```

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 4, 2000 this sequence version replaced g1:7259782.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

## Project Information

Center project name: L8577

Center clone name: 635\_O\_11

## Summary Statistics

Sequencing vector: M13: M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 168929 bases at least Q40  
 Consensus quality: 172827 bases at least Q20  
 Consensus quality: 174151 bases at least Q20  
 Insert size: 176000; agarose-fp  
 Insert size: 174877; sum-of-contigs  
 Quality coverage: 4.5 in Q20 bases; agarose-fp  
 Quality coverage: 4.6 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1      2997: contig of 2997 bp in length
*
*      2998 3097: gap of 100 bp
*
*      3098 5606: contig of 2509 bp in length
*
*      5607 5706: gap of 100 bp
*
*      5707 9209: contig of 3503 bp in length
*
*      9210 9309: gap of 100 bp
*
*      9310 13614: contig of 4305 bp in length
*
*      13615 13714: gap of 100 bp
*
*      13715 20639: contig of 6925 bp in length
*
*      20640 20739: gap of 100 bp
*
*      20740 26545: contig of 5806 bp in length
*
*      26546 26645: gap of 100 bp
*
*      26646 37068: contig of 10423 bp in length
*
*      37069 37168: gap of 100 bp
*
*      37169 48083: contig of 10915 bp in length
*
*      48084 48183: gap of 100 bp
*
*      48184 59015: contig of 10832 bp in length
*
*      59016 59115: gap of 100 bp
*
*      59116 70420: contig of 11305 bp in length
*
*      70421 70520: gap of 100 bp
*
*      70521 80466: contig of 9946 bp in length
*
*      80467 80566: gap of 100 bp
*
*      80567 92325: contig of 11759 bp in length
*
*      92326 92425: gap of 100 bp
*
*      92426 110572: contig of 18147 bp in length
*
*      110573 110672: gap of 100 bp
*
*      110673 131898: contig of 21226 bp in length
*
*      131899 131998: gap of 100 bp
*
*      131999 176277: contig of 44279 bp in length.

```

## FEATURES

## source

```

1..176277
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-635011"
/clone_lib="RPC1-11 Human Male BAC"
1..2997
misc_feature
/feature="assembly_fragment"
3098..5606
misc_feature

```

```

/misc_feature
/feature="assembly_fragment"
5707..9209
vector_side:left"
/misc_feature
/feature="assembly_fragment"
9310..13614
/misc_feature
/feature="assembly_fragment"
13715..20639
/misc_feature
/feature="assembly_fragment"
20740..26545
/misc_feature
/feature="assembly_fragment"
clone_end:r7
vector_side:right"
/misc_feature
/feature="assembly_fragment"
26646..37068
/misc_feature
/feature="assembly_fragment"
37169..48083
/misc_feature
/feature="assembly_fragment"
48184..59015
/misc_feature
/feature="assembly_fragment"
59116..70420
/misc_feature
/feature="assembly_fragment"
70521..80466
/misc_feature
/feature="assembly_fragment"
80567..92325
/misc_feature
/feature="assembly_fragment"
92426..110572
/misc_feature
/feature="assembly_fragment"
110673..131898
/misc_feature
/feature="assembly_fragment"
131999..176277
/feature="assembly_fragment"
/feature="assembly_fragment"

BASE COUNT  54493 a 33357 c 33254 g 53772 t 1401 others
ORIGIN

```

```

Query Match      99.8%; Score 940.4; DB 2; Length 176277;
Best Local Similarity 99.9%; Pred. No. 1.6e-258;
Matches 941; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 atgaattgggttaagaagcagcatcacagaggttatctctcgtggttcacagatcga 60
|||||
DB 154174 ATGAATGGGTAAAGACGACATACAGAGGATTATCTGGTTTCACAGATCGA 154115
|||||
QY      61 ccttgctgagcttccacctctgtgtctcttgatcttctacactgtgacatctt 120
|||||
DB 154114 CCTTGCTGAGATTTCACATCTCTTGCTGCTCTTGATTTCTTACACTGTGACATCTT 154055
|||||
QY      121 ggcacatctgacatattcttagtgcgctgtgacacacacaaattcatacccatglat 180
|||||
DB 154054 GGCACATCTGACCATTTACTAGTGCACGCCGTGACACCAACTTCATACCCCATGTTAT 153995
|||||
QY      181 tttttcttaccacatctacactctctgactctgttcttgaacacacatgtaagtcacaaa 240
|||||
DB 153994 TTTTTCCTTACCAACTATCACTCTCTGCAATCTTTGTATACCAACATGTACAGTCCACAA 153935
|||||
QY      241 atgcctagtaaatlctgacgacatcagaaagaatcaagtlatactgctgtgtgtaagccag 300
|||||
DB 153934 ATGCCTAGTAATTTATMGACGATCAGGAAGAATATCAGTTATCGTGCGCTGTATACCCAG 153875
|||||
QY      301 ctttccatactctggtccttgggggactgtaatactctctcctggtgcgtatgtccctt 360
|||||
DB 153874 CTTTTCATATTTCTGGCCTTGGGGCTACTGAAATCTTCTCTGCGCTCATGTCCTTT 153815
|||||
QY      361 gataagttgtgactatctgtgctcttccatctcctcaagtatatgacacagagacac 420
|||||
DB 153814 GATAGGTTGTGATGATTTGTGGCCTCTCCATCTCAGTTATATGACACGAGACATC 153755
|||||
QY      421 tgcctccagttgacgacgcacccctgggttaactggttttagtaactcagtggtgtctc 480
|||||
DB 153754 TGCCCTCCAGTTGGACGCTGCATCTGGGTTACTGTTTAGTAACACTGATGCTGCTT 153695
|||||
QY      481 accctgactctccagctgcacactctgtgacccctatgtgatagatcaactttctgtgaa 540
|||||

```



Db 153694 ACCCTGACTCTCCAGCTGCACTCTGTGACCCCTATGTGATGATCACTTCTGTGAA 153635  
 Oy 541 gtccctgacgtcctaagttatctgtgttgagacaagcaagagcgtgaactatc 600  
 Db 153634 GTCCCTGACAGCTCAAGTTATCTGTGTGAGACACAGCAAAAGAGCGTCAACTATTC 153575  
 Oy 601 ctgtcagtgagctctccatcaataacccccgacacccatccatatacatgcttt 660  
 Db 153574 CTGTGACATGAGCTCTTCATCATATATACCCCTGACACATCATCTTATATCATATCTTTT 153515  
 Oy 661 attgtccgagcagatltgagatacagtcgctggaagtcgacaaagacatttgagaca 720  
 Db 153514 ATTGTCCAGCAGCATTTGAGATACAGTCTGTGAAGTCGACAAAGACATTTGGGACA 153455  
 Oy 721 tgtgttcccatctaatgtgtgtgtcttttttataagacgcgtctgtgtactgt 780  
 Db 153454 TGTGTTTCCCATCTAATGTGTGTCTTTTATGATACGCGTCTGTGTACTGTG 153395  
 Oy 781 caaccacgtccgacgtcccaagacagcaagaaagatgtgtctctctctatggaac 840  
 Db 153394 CAACCACTTGCGCCAGCTCCAGAGCAAGCAAGAAAGATGTTTCTCTCTATGGAATC 153335  
 Oy 841 attgacacacgtcgtgaatcccccttatatacacttaggaacaagaggtgaaggaagc 900  
 Db 153334 ATTGCACCCAGTCTGAATCCCTTATATATACACTTAGGAACAAGAGGTAAAGAGAGGC 153275  
 Oy 901 tttaaaagttggtgtgcaagagctcttcttaatcaagaataa 942  
 Db 153274 TTTTAAAGTTGGTTGTGCAAGAGTCTTCTTAATCAAGAAATAA 153233

RESULT 3  
 HSA302593 1074 bp DNA PRI 06-JAN-2001  
 LOCUS Homo sapiens 6M1-10\*02 gene for olfactory receptor, cell 11ne  
 DEFINITION KR3598.  
 ACCESSION AJ302593  
 VERSION AJ302593.1 GI:12054410  
 KEYWORDS 6M1-10\*02 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J.,  
 Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 (in) Kasahara, M. (Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)  
 REFERENCE 2 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
 Younger, R. M. and Beck, S.  
 Polymorphisms in olfactory receptor genes: a cautionary note  
 Unpublished  
 3 (bases 1 to 1074)  
 REFERENCE Younger, R. M., Amado, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
 Forbes, S. A., Horton, R., Milne, S., Mungall, A. J., Trowsdale, J.,  
 Volz, A., Ziegler, A. and Beck, S.  
 Characterisation of clustered MHC-linked Olfactory Receptor Genes  
 in Human and Mouse  
 Unpublished  
 4 (bases 1 to 1074)  
 REFERENCE Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.,  
 Ziegler, A.  
 MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 Unpublished  
 5 (bases 1 to 1074)  
 REFERENCE Ehlers, A.  
 Direct Submission

JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY  
 FEATURES  
 source Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="KR3598"  
 1..1074  
 /gene="6M1-10\*02"  
 1..1074  
 /gene="6M1-10\*02"  
 /codon\_start=1  
 /product="olfactory receptor"  
 /protein\_id="CAC20513.1"  
 /db\_xref="GI:12054411"  
 /translation="MNMYNKSVPQEFLLVSPDQPMLEIPPFVMEFSYILTFIGNLT  
 ILVSHVDFKLHPMPFELSNLSLDLCYPTSTVPOMLVNCTRKYVSGGCAOLF  
 IFALGSECEILLAVMCFEDRVALCRPLHSYIMHQRCEPOLAAASMSGFSNVSLOS  
 TWLKMPLGHRKVDHFEVPAFLKUSVDPTNEMAEFTSYFLIPVTLILSY  
 AFIVQAVLRISAGRRKRAFGTSGSHLIVSLFYGTALISMYLQPPSPSSDKRWKWSL  
 FCGIILAPMLNPLIYTLRNKEVKEAFKRLVAKSLINOETIRNMQISFAKDVLTFLTNF  
 SASCPFIETIENYCNLPQRKP"  
 BASE COUNT 277 a 248 c 203 g 346 t  
 ORIGIN  
 Query Match 77.3%; Score 727.8; DB 9; Length 1074;  
 Best Local Similarity 86.4%; Pred. No. 9.7e-198;  
 Matches 815; Conservative 0; Mismatches 127; Indels 1; Gaps 1;  
 Oy 1 atgaattggttaaatgacagcatcacagaagattatctgctggtgtctcagatcga 60  
 Db 1 ATGAATTGGGTAAATAAGAGTGTCACAGAGTTCATCTGTAGTTTCTCAGATCAA 60  
 Oy 61 ccttgctggagttccacactcctctgtgtgtctcttgattctctaacctgtgacatctt 120  
 Db 61 CCATGGCTAGAGATTCCACCCCTTGTGTGATGTTGTTTCTTATATCTTGACAAATCTTT 120  
 Oy 121 ggaactgtaccattatctctagtcagcgcctgacacaaactcacaacccccatgtat 180  
 Db 121 GGCATCTGACATATATTTCTTGTGCACATGTGGATTTCAACTCCACAGCCCTATGTAC 180  
 Oy 181 ttttcttccaactatcacctcctgtgactctgttcttaccacacatgtaacagtcacaa 240  
 Db 181 TTTTCTTTAGCAATCTCTCACTCGACCTTGACTATACCAAGTACAGTCCACAA 240  
 Oy 241 atgctagtaaatltatgcagcatcaggaagtaacatcagttacgtggtcgtgtacccag 300  
 Db 241 ATGCTGTAAACATATGCAACACAGCAAGAAATACATGATGATGCGCTGTGCGCCAG 300  
 Oy 301 ctttcataatcttgcccttgagggtgactgtaatactctcctcgccgtcaatgctct 360  
 Db 301 CTTTTCATTTTCTTGCGCTTGAGGTTCCACAGATATGTTCTCTGCGCTCATGTGCTTT 360  
 Oy 361 gataagttgtagctattgttcgacctcctacactacgaattacatgacacagaagtc 420  
 Db 361 GATAGATTGTGATGATTTGTGGCCCTCCATTAATCAATATATGATGACACAGAGGCTC 420  
 Oy 421 tgcctccagttggcagcgcatccctgggttactggttttagtaactcagtggtgtgtct 480  
 Db 421 TGCTTCCAGTTGGCAGCGTCATCGGATTAATGCTTTAGCAATTAATTAACAGTCC 480  
 Oy 481 accctgacctccagctcagctcgtggaacccctatgtatgatatcattctcgtgga 540  
 Db 481 ACCTGACACTTAAGATGCACTGTGTGTGTCACAAAGATGATGATCTTCTTGTGTGA 540  
 Oy 541 gtccctgacgtcctaagttatctgtgttgagacaagcaagagcgtgaactatc 600  
 Db 541 GTCCCTGACGTCTCAAGTTGTCTGTGTGACACAAAGCAAGAGCGTGAACATATTC 600  
 Oy 601 ctgtcagtgagctcttccatcaataaccccgacacatcattcattacatgcttt 660



QY 721 tgggtcccaactaatgtgtgtgtctttttatagtaacagccgtctgtgtacgtg 780  
DB 721 TGTGGCTCCCAATCAATGTGTGTGTCTTTTATATGTATGATCTCATCTGACCTG 780  
QY 781 caaccacctgcgcagctcccaaggacaaagaatgtgtctctctctatgtgaac 840  
DB 781 CAACCACTTCCAGCTCCAGCTCCAGGAGGAAAGATGTCTCTCTCTGTGATC 840  
QY 841 attcaccacatgcgtatcccttatatactactaggaacaaagaagtaagaagc 900  
DB 841 ATTGACCCATGCTGAATCCCTTATATATACCTTAGAACAAAGAGTAAGGAGCC 900  
QY 901 tttaaaagttgtgtgc-aagagctctctatcaagaataa 942  
DB 901 TTTAAAGTTGTGTGCAAGAGCTCTTATATCAAGAAATTA 943  
RESULT 5  
HSA302585 1074 bp DNA PRI 06-JAN-2001  
LOCUS Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line  
DEFINITION BM19.7.  
ACCESSION AJ302585 GI:12054394  
VERSION 6M1-10\*01 gene; olfactory receptor.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1074)  
AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(in) Kasahara, M. (Ed.);  
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
FUNCTION: 110-130;  
Springer-Verlag, Tokyo, Japan (2000)  
REFERENCE 2 (bases 1 to 1074)  
AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
Younger, R. M. and Beck, S.  
TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
Unpublished  
3 (bases 1 to 1074)  
AUTHORS Younger, R. M., Amadou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
Forbes, S. A., Horton, R., Milne, S., Mungall, A. J., Trowsdale, J.,  
Volz, A., Ziegler, A. and Beck, S.  
TITLE Characterisation of clustered MHC-linked olfactory Receptor Genes  
in Human and Mouse  
Unpublished  
4 (bases 1 to 1074)  
AUTHORS Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.  
and Ziegler, A.  
TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes  
Unpublished  
5 (bases 1 to 1074)  
AUTHORS Ehlers, A.  
TITLE Direct Submission  
Submitted (06-NOV-2000) Ehlers, A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY  
FEATURES  
source Location/Qualifiers  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="BM19.7"  
1..1074  
/gene="6M1-10\*01"  
1..1074  
/gene="6M1-10\*01"  
/codon\_start=1  
CDS

/product="Olfactory receptor"  
/protein\_id="CAC20505.1"  
/db\_xref="GI:12054395"  
/translation="MMWNKSVPOEFLILVFSDDPMLFIPFVMEFLSYILITIGNL  
IIVSHVDPLKHPMPYFELSNLSLDLCYTSTVPMILVNCNRKVISGGVVAOLF  
IFLALGSETECLLVAMCPDFREVAICRDLPHS1INHORICEPOLAAASVSGSVLS  
TWTLKPLCGREVDHFECEPALKTSCVPTTNEAEFLFISVFLIPVTLILISY  
AFIVQAVLR10SARGORKAREGTSNHLIVSLFVGPALISWTLQPPSPSRKQKMSVL  
FCGIIAEMLNPLIYTLNKEVKEAFKRLVAKSLNQLRNMQMSFAKDYVLTITNF  
SASCPITVITLTCNLPORKEP"  
BASE COUNT 278 a 248 c 202 g 346 t  
ORIGIN  
Query Match 77.1%; Score 726.2; DB 9; Length 1074;  
Best Local Similarity 86.3%; Pred. No. 2.8e-197;  
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;  
QY 1 atgaattggaataagacacatcatagagagattatctgtgtgttcagatcga 60  
DB 1 ATGAATGGGTAATAAGAGTGTCCACAGAGGATCTGTGATTCTCAGATCA 60  
QY 61 ccttgctgaggttccactccctgtgtgtctctgtatctcttaccatctgacatctt 120  
DB 61 CCATGCTAGAGATTCACACCCCTTGTGATGTTCTGTTCTCTATATCTGACATCTTT 120  
QY 61 ggaactctgacattatcttagtgcagcctgtgagacccaactcaaccatgtat 180  
DB 121 GGCAATCTGACAAATATCTTGTGTACATGTGGATTTCAACCTCACACCCCTTATGTC 180  
QY 181 tttcttaccacatctactcctgtgtatctcttcttaccacagatgtaagctccaca 240  
DB 181 TTTTCTTCTTACCAATCTCTCACTCTGTGACCTTCTCTTATACCAAGATCACTTCCACAA 240  
QY 241 atgcagtaaatlttagcagacatcaggaagtaacatgaltacgtggtctgtgacccag 300  
DB 241 ATGCTGGTAACATATGACACACACAGGAAGTAATCAGTTAAGTGGTGTGGCCAG 300  
QY 301 ctttcataatctgtgcttgggggtctacgaatactcttccctgtgcgtatgtccctt 360  
DB 301 CTTTTCATTTTCCCTGCGCTTGGGTTCCACAGATGTCCTTCCTGCGCTCATGTGCTTT 360  
QY 361 gataagttgtatgctattgtgtggtctctccatctcaagtatcatcacaagagact 420  
DB 361 GATAGCTTTTATGCTATTGTTGCGCTCTCCATCTCAATATATATGACACAGAGCTC 420  
QY 421 tgcctccagttgagcagccgacatccctgggttactgtgtttagtaactcagttgtgtct 480  
DB 421 TGTCTCCAGTTGGCAGCTGCAGTCCTGATTTAGTGGCTTTAGCAATTCAGATTACAGTCC 480  
QY 481 accctgactctcagcgtccactctgtgaaccctatgtgatagtaactcttctctgga 540  
DB 481 ACCCTGACACTTAAGATGACCTGTGTGTGTCACAAAGAAAGTGAATCCTTCTTGTGTA 540  
QY 541 gttccctgacgtctaaatctctgtgtgtttagaacaagaagaagagctgaactatct 600  
DB 541 GTTCCCTGCTGTCTAAAGTGTCTGTGTGACACACAGCAAAATGAGGCTTAACCTATTTC 600  
QY 601 ctgtcagtgagctctccatcctaataccctgacacatccatcctatatacatgcttt 660  
DB 601 TTTCATCAGTGTGCTATCTCTTCTTATACCCGACACATCATTATATCGATCTTTT 660  
QY 661 attgtcgaagcagatlttagatatacgtctgtgaagtcgacaaaagaatctgtggaca 720  
DB 661 ATTGTCCAAGCAGTCTTGAAGATCCAGTCTGTGAAGTCAAGCAAGGCAATTTGGGACA 720  
QY 721 tgggttccatctaatgtgtgtgtctttttatagtaacagccgtctgtgtgtacgtg 780  
DB 721 TGTGGCTCCCAATCAATGTGTGTGTCTTTTATATGTATGATCTCATCTGACCTG 780  
QY 781 caaccacctgcgcagctcccaaggacaaagaatgtgtctctctctatgtgaac 840  
DB 781 CAACCACTTCCAGCTCCAGCTCCAGGAGGAAAGATGTCTCTCTGTGATC 840

QY 841 attgcaccatcgtgaatcccttatatacacttaggaacaaggagtaagaagc 900  
 |||||||  
 Db 841 ATTGCACCATGCTGAATCCCTTATATATACACTTAGGAACAAGAGTAAGAAGCC 900  
 QY 901 tttaaaggctgtgtgc-aaggctcttctaatacaagaata 942  
 |||||||  
 Db 901 TTTAAAGGCTGTGCAAGAGCTCTTATCAAGAAATA 943  
 RESULT 6  
 HSA302586 1074 bp DNA PRI 06-JAN-2001  
 LOCUS Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line LG2.  
 DEFINITION  
 ACCESSION AJ302586.1 GI:12054396  
 VERSION 6M1-10\*01 gene; olfactory receptor.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J.,  
 Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 (in) Kasahara, M. (Ed.);  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)  
 REFERENCE 2 (bases 1 to 1074)  
 Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
 Younger, R. M. and Beck, S.  
 Polymorphisms in olfactory receptor genes: a cautionary note  
 Unpublished  
 3 (bases 1 to 1074)  
 Younger, R. M., Amadou, C., Bethel, G., Ehlers, A., Fischer, L., Lindahl, K.,  
 Forbes, S. A., Horton, R., Milne, S., Mungall, A. J., Trowsdale, J.,  
 Volz, A., Ziegler, A. and Beck, S.  
 Characterisation of clustered MHC-linked olfactory Receptor Genes  
 in Human and Mouse  
 Unpublished  
 4 (bases 1 to 1074)  
 Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.,  
 and Ziegler, A.  
 MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 Unpublished  
 5 (bases 1 to 1074)  
 Ehlers, A.  
 Direct Submission  
 Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY  
 FEATURES  
 SOURCE  
 Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="LG2"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /codon\_start=1  
 /product="olfactory receptor"  
 /protein\_id="CAC20506.1"  
 /db\_xref="GI:12054397"  
 /translation="MNWYKSVPOEFLVSDPMLIEPFEVLEFSYILITGNLT  
 ILVSHDFKHTPMYFELSNLSLDLCYTSYVPMKLVNVCNTRKYSVGGCAOLF  
 IFALGSTECILLAVKCFDRVAICRPLHYSITNHQRCRFLAASWISGFSNVSLOS  
 TWLMPKCGHKEVDHFCFCEVPAALKLSCVDTTANEALEFISVLFILPTLILISY  
 AFIVQAVLRIOSAGORAKFCGSHLIVSLFYGTAISMVLOPSPSSKDRGKRWSL

Query Match 77.1%; Score 726.2; DB 9; Length 1074;  
 Best Local Similarity 86.3%; Pred. No. 2.8e-197;  
 Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;  
 ORIGIN  
 ECIIAPMLNPLIYLRNKEVKEAFKRLVAKSLNNOEIRNNOMISFAKDTLVLYTLNF  
 SASCPHFEVITTEYCNLPORKEP"  
 BASE COUNT 278 a 248 c 202 g 346 t  
 QY 1 atgaattggttaagacagcatcacacagaggttatctctcgtggttcccaatgca 60  
 |||||||  
 Db 1 ATGAATGGGTAATAAGAGTGTCCACAGGAGTTCATCTGTAGTTTCCACATACA 60  
 QY 61 ccttgagtgagttcccaactctcttggtctcttgatcttcaacttgagaccatt 120  
 |||||||  
 Db 61 CCATGGCTAGAAATTCACACCTTTGTGAGAGTTTCTTCTTCTATATCTTGCAATCTTT 120  
 QY 121 ggcacatctgaccattatcttagtgcacgccttgagacacccaactcatalccca 180  
 |||||||  
 Db 121 GCAATCTGACATATATTTCTGTGCACATGTGATTCAACTGCACACCCCTATGTAC 180  
 QY 181 tttttcttaacaatcatcatctcttgatctcttgtaaccaatgtaagttccaca 240  
 |||||||  
 Db 181 TTTTTCCTTAGCAATCTCTCACTCTGACCTTCTCTATACCAAGTACAGTTCCACAA 240  
 QY 241 atgcctagaatttatgcagcatcaggaagaatcatcagttacgtgagctgttagccag 300  
 |||||||  
 Db 241 ATGCTGGTAACATATGACACACCGAAGAAATACAGTTATGTGTGCTGTGTGCCAG 300  
 QY 301 ctcttcatactcttgaccttgagggtctactgaatactctctccttgccgtcatct 360  
 |||||||  
 Db 301 CTTTCATTTTCTGCGCTTGCGCTTGCCACAGATGCTTCTGCGCTCATGTGCTTT 360  
 QY 361 gataagttgtatgactatctgtcgccctcctcaactcaagttatcaatgacacagagctc 420  
 |||||||  
 Db 361 GATAGGTTGTAGCATATTTGTGCGCTCTCCATCTCATATATCATGACACGAGGCTC 420  
 QY 421 tgcctcagatggagccagccatccctgggttactggttttagaactcaagtgagttgct 480  
 |||||||  
 Db 421 TCTTTCACATTTGGACGCTGACATCTGACATTAAGTGGCTTTAGCAATTCAGATTTACAGTCC 480  
 QY 481 accctgacctcctcagctcctcactcgttgacccctatgtgatagatacaacttctcgtgaa 540  
 |||||||  
 Db 481 ACCTGGACACTTAACATATGACATGCTGTGTACCAAGATGATGACATCTTCTGTGTA 540  
 QY 541 gtccctgcacgtcctcaagtatctctgtgttgagacacagcaatgagctgtaactatc 600  
 |||||||  
 Db 541 GTCCCTGCTCTCTCAAGTTGTCTGTGTGACACACAGCAAAATGAGGCTGTAACATAATTC 600  
 QY 601 ctgtcagtgagctcttcacatcattacccctgagacatcattcattatcattatcatt 660  
 |||||||  
 Db 601 TTTCATCAGTGTCTTATCTTCTTATATACCGTGACATCATCTTATATCTGATCTTTT 660  
 QY 661 attgtccgagcagtaattgagatatacagctcgtgaaggtcgacaaagacatttgagaca 720  
 |||||||  
 Db 661 ATTGTCCAAGCAGTGTGAGATTCACAGTGTGTAAGGTCACAGAAAGCAATTTGGACA 720  
 QY 721 tgtgttcccatcattatgtgtgtctctttttatagtaacagccgtctcgtgtactct 780  
 |||||||  
 Db 721 TGTGTCTCCATCTAATGTGTGTCTCTTATGTGTACAGCTATCTCCATGTACCTG 780  
 QY 781 caacacacctggccagctcctcaaggaacaaagaaatgagttctctcttcattagaaatc 840  
 |||||||  
 Db 781 CAACACACCTTCACCCAGCTCCAAAGACCGGGAAAGATGTTCTCTCTGTGTGAATC 840  
 QY 841 attgcaccatcgtgaatcccttatatacacttaggaacaaggagtaagaagc 900  
 |||||||  
 Db 841 ATTGCACCATGCTGAATCCCTTATATATACACTTAGGAACAAGAGTAAGAAGCC 900  
 QY 901 tttaaaggctgtgtgc-aaggctcttctaatacaagaata 942  
 |||||||  
 Db 901 TTTAAAGGCTGTGCAAGAGCTCTTATCAAGAAATA 943

RESULT 7  
HSA302587 1074 bp DNA PRI 06-JAN-2001  
LOCUS Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line  
DEFINITION H21CL.  
ACCESSION AJ302587  
VERSION 1 GI:12054398  
KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1074)  
Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J.,  
Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(in) Kasahara, M. (Ed.):  
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
FUNCTION: 110-130;  
Springer-Verlag, Tokyo, Japan (2000)  
REFERENCE 2 (bases 1 to 1074)  
Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A.,  
Younger, R.M. and Beck, S.  
Polymorphisms in olfactory receptor genes: a cautionary note  
Unpublished  
3 (bases 1 to 1074)  
Younger, R.M., Amadou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
Forbes, S.A., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J.,  
Volz, A., Ziegler, A. and Beck, S.  
Characterisation of clustered MHC-linked Olfactory Receptor Genes  
in Human and Mouse  
Unpublished  
4 (bases 1 to 1074)  
Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.  
and Ziegler, A.  
MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes  
Unpublished  
5 (bases 1 to 1074)  
Ehlers, A.  
Direct Submission  
Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY  
FEATURES  
source Location/Qualifiers  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="H21CL"  
1..1074  
/gene="6M1-10\*01"  
1..1074  
/gene="6M1-10\*01"  
/product="Olfactory receptor"  
/protein\_id="CAC20507.1"  
/db\_xref="GI:12054398"  
/translation="MNMVNSVPOEFLIVSDPMLFIPPMVLFVYLITFCNLT  
ILVSHVDFKLTHTPRFFLSNLSLDLCYTTSTVQMLVNTCNKTVSYGCAQLF  
IFLAGSTECFLAWCFDFRVAICRPVHSTVHRCFOIAASWISGFSNVLOS  
TWTLMPLCGHEDVDFCEVPALIKSCVDITANAEALFFISVFLIPVLLIISY  
AFIVQAVRIQSAEGORAKFGTSGSHLIVSLFTGTALISMYLQPPSPSKDGKAVSL  
FCGIIAPMLNPLIYLRNKEVKAERFLVAKSLNQEIRNMQIMISFANDVTVLTINE  
SASCPITFVITTEYCNLPQRKPP"  
BASE COUNT 278 a 248 c 202 g 346 t  
ORIGIN  
Query Match 77.1%; Score 726.2; DB 9; Length 1074;

Best Local Similarity 86.3%; Pred. No. 2,8e-197;  
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;  
QY 1 atgaattggttaaatagacagacatcacagagattatctgctggtgttctcagatcga 60  
DB 1 ATGAATTGGTGAATAGAGTGTCCACAGAGATTGATCTGTAGTTTTCAGATCAA 60  
QY 61 ccttgctggaatttcccaactcctcttggtctctcttgattcttacaactggtccatt 120  
DB 61 CCATGGCTAAGAAATTCACACCTTTGTGATGTTGTTGCTTATATCTTACAAATCTTT 120  
QY 121 ggcacatgacattatctagtgcaagccctgagacacacattacatccccatgatt 180  
DB 121 GGCATCTGACAAATATTTCTTGTGTCACATGTGGATTTCAACTCCACACCCCTATGTAC 180  
QY 181 ttttcttacaactatcacactcctcgtgacttctgtttacacacatgtaagtcacaa 240  
DB 181 TTTTCTTACGAATCTCCTGACCTCGACCTTGGTATACCAAGATGACATTCACAA 240  
QY 241 atgctagtaattatgacgacatcaggaagaatcagttatcgtgctgtagccag 300  
DB 241 ATGCTGTAAACATATGCAACACAGGAAGTAATCAGTTATGTGGCTGTGGCCAG 300  
QY 301 ctttcaatactctgagccttgagggtgactgaaatactctcctgagcgtcatgtcctt 360  
DB 301 CTTTCAATTTTCTGCGCTTGAGGCTTCACAGATGCTTCTGCGCTCATGTGCTTT 360  
QY 361 gataagttgtatgactatctgctgagcctcctcattactacagtatacagacagagatc 420  
DB 361 GATAGTTGTGTAGCTATTTGTGGCCTCTCCATTAATATATCATGACACAGAGGCTC 420  
QY 421 tgccttcaattgagcagcagcctcctggtgtactggtttgaatcctcagatggtgtgtc 480  
DB 421 TGCCTTCAAGTTGACAGCTCATCTGAGTTAGTGCTTGAATTCATTAATTAACAGTTC 480  
QY 481 accctgacctcctcagctgacactctgagacctatgatatagatcaactctctgtga 540  
DB 481 ACCTGACACTTAAGATGACACTGTGTGTCACAAAGATGTGATCTTCTGTGTGA 540  
QY 541 gtccctgacactgctcaagttatctctgtgtgagacacacagcaaatgagctgaactatc 600  
DB 541 GTCCCTGCTCTGCTCAAGTTGTGTGTCAGACAAACAGCAAAATGAGCTGAACATATTC 600  
QY 601 ctgttcagtagctcttccatctcaataacccctgagacctatccttatcatatgcttt 660  
DB 601 TTTATGAGTGTCTATCTCTTCTTAATACCCGTAACCTATCTTAATGATGATGCTTTT 660  
QY 661 attgtccgagcagatctgagagatagatagctgctgagagctgagacaaagacattggaga 720  
DB 661 ATTGTCCAGCAGCTGTGAGATCCAGTCTGCTGAAGGTCAACGAAGGACATTTGGGACA 720  
QY 721 tctgtgtccatctaatctgtgtgtctcttttattagtagacagcgctcctgtagctgt 780  
DB 721 TGTGCTCCCATCTAATTTGTGTGTCATTTTATGTAACGCTATCTCATGTGCTGTG 780  
QY 781 caaccacctgcccagctccaaagacaaagaaatgatttcttcttcatggaatc 840  
DB 781 CAACCACTTCCACCCGCTCAAGACCGGGAAGATGATGTTCTCTCTGTGTGAATC 840  
QY 841 attgacccatctgtaactcccttatatactaggaagaaagaaagaaagaaagc 900  
DB 841 ATTGACCCATGCTGAATCCCTTATATATATACCTTAGGACAAAGAGTAAGGAAGCC 900  
QY 901 tttaaaagctgtgtgc-aagagcttcttaacaaagaaataa 942  
DB 901 TTTAAAGCTGTGTTCAAAAGCTTCTTAATCAAGAAATAA 943  
RESULT 8  
HSA302588 1074 bp DNA PRI 06-JAN-2001  
LOCUS Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line WT51.  
ACCESSION AJ302588

VERSION AJ302588.1 GI:12054400  
 KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J.,  
 Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 JOURNAL (in) Kasahara, M. (Ed.);  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)  
 REFERENCE 2 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A.,  
 Younger, R.M. and Beck, S.  
 TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1074)  
 AUTHORS Younger, R.M., Amdou, C., Bethel, G., Ehlers, A., Fischer, L., K.,  
 Forbes, S.A., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J.,  
 Volz, A., Ziegler, A. and Beck, S.  
 TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
 in Human and Mouse  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 1074)  
 AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.  
 and Ziegler, A.  
 TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 JOURNAL Unpublished  
 REFERENCE 5 (bases 1 to 1074)  
 AUTHORS Ehlers, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-2000) Ehlers, A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY  
 FEATURES  
 source location/Qualifiers  
 1..1074 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="WT51"  
 1..1074 /gene="6M1-10\*01"  
 1..1074 /gene="6M1-10\*01"  
 /gene="6M1-10\*01"  
 /product="olfactory receptor"  
 /codon\_start=1  
 /protein\_id="CAC20508.1"  
 /db\_xref="GI:12054401"  
 /translation="MNWYKSVPEFILLVSDPWLIEPPWFLESLYILTINSLT  
 IIVSHDFKLHPYFPLSLNLSLDLYTSTVQMLVNTKRVISYGGCAQLF  
 IFLAGSTECFLLAVMCFDRAICRPLHYSIMHQRCLFQIAASWISGNSVLS  
 TWTLKMLPGCKHEVDHFCFCEVPALIKSCVDPTANAELEFISVFLIPYLLIISY  
 AFIYOAVLRIOASBGRGFGSHLIVSVLFGTATISMTLOPSPSKRGKWSL  
 FCGILAPMLPLIYLRNKEVKEARKRLVANSLSLNOELRNQMISFADYVLTILTNF  
 SASCPFEVITTEINCNLPORRKP"  
 BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN  
 Query Match 77 1%; Score 726.2; DB 9; Length 1074;  
 Best Local Similarity 86.3%; Pred. No. 2.8e-197;  
 Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;  
 Oy 1 atgaattggtaaatgacacatcatatctctctgaggttccaaagtcga 60  
 |||||  
 Db 1 ATGAAATGGGTAAATGAAGGTGTCCACAGGAGTTTCATCTGTAGATTTCACAGTCAA 60  
 |||||  
 Oy 61 cctgtgaggttccactcctgtgtctctctgattcttaacactgtgacatcctt 120  
 |||||

|||||  
 Db 61 CCATGGCTAGAGATTCACACCTTGGATGTTCTGTTTCCATATCTTGACAACTTT 120  
 |||||  
 Oy 121 ggaacatcgaacatattctagtgtacgctcgtgacacaaactataccccaatgat 180  
 |||||  
 Db 121 GGAATCTGACAAATATCTGTGTGTCATGTGGATTTCAAACTCCACACCCCTATGTAC 180  
 |||||  
 Oy 181 ttcttcttaacatataccctcgtgactctgttgaaccacagtcagtcacaa 240  
 |||||  
 Db 181 TTTTCTTAGCATATCTCACCTCGGACCTTGGCTATACCAAGTACAGTTCCACAA 240  
 |||||  
 Oy 241 atgctagaatattatgcagacatccatcgtgtaactgattatcgtgtgtgacag 300  
 |||||  
 Db 241 ATGCTGTAAACATATGACAAACACAGGAAGTAAATGATGTTGCTGTGGCCAG 300  
 |||||  
 Oy 301 ctttcatattctcgtccttggggcctactgaatatctctcctggcgtatgcctt 360  
 |||||  
 Db 301 CTTTTCATTTTCCCTGGCCCTTGGGTTCCAGAAATGCTTCTCGCCGTCATGTCCTT 360  
 |||||  
 Oy 361 gataggttgtagctattcgtcggcctcccatcactcaagtatacgtacacagagac 420  
 |||||  
 Db 361 GATAGGTTGTAGCTATTTGTGCGCTCTCCATTACTCATTTATTCACACAGAGCTC 420  
 |||||  
 Oy 421 tgcctccatcgtgacgcacatccctgtggtactggtttagtaactcagtggtgtct 480  
 |||||  
 Db 421 TGCTTCAGTTGGCAGCTGCATCTCGATTAGTGCGCTTAAAGCAATTCAGATTACAGTCC 480  
 |||||  
 Oy 481 accctgaactcctcagctcgaactcgttgacccctatgtagatagatcaacttctctg 540  
 |||||  
 Db 481 ACCTGGACACTTAAGATGCGACACTGTGTCACAAAGATGATGATCTTCTGTGAA 540  
 |||||  
 Oy 541 gtcccgacactgctcaagatctctgtgttgagacacagcaatgagctgaactatc 600  
 |||||  
 Db 541 GTCCCTGCTCTGCTCAAGTTGCTCTGTGTGACACACAGCAATGAGCTGAATATTC 600  
 |||||  
 Oy 601 ctgtcagtgagctctctccatcaataccctcgtgacacatccatcatatagcttct 660  
 |||||  
 Db 601 TTCATGAGTGTGCTATTCCTTGAATATACCCGAGACATCATCTTATATCGATGCTTT 660  
 |||||  
 Oy 661 attgtcgaagcagatattgagagatagatcgtcgtgaagtcgcagaaagaacattgg 720  
 |||||  
 Db 661 ATTGTCCAGACAGTGTGTAATTCACAGTCTGTAAGGTCACAAAGGCAATTTGGGACA 720  
 |||||  
 Oy 721 tctgtctccatcaattatctgtgtctcttttaataagtagcagcgctcgtgtgac 780  
 |||||  
 Db 721 TGTGCTCCCATCTAATTTGTGTGTACTTTTATGATGACACTATCTCATGTACTG 780  
 |||||  
 Oy 781 caacacactcgcgcac 840  
 |||||  
 Db 781 CAACACACTTCAACCCAGCTCCAAAGACCGGGGAAGATGTTCTCTCTGTGTGAATC 840  
 |||||  
 Oy 841 attgacccatcgtgaatcccttataatatacacttagaacaagaagtagaagaagc 900  
 |||||  
 Db 841 ATTGACCCATCTGTGAATCCCTTATATATACACTTAGAACAAGAGTAAAGAACCC 900  
 |||||  
 Oy 901 tttaaaagttgtgtgc-aagagctcttaacttaagaagaataa 942  
 |||||  
 Db 901 TTTAAAGTTGTGTGCAAGAGCTCTTTTATCAAGAAATAA 943  
 |||||  
 RESULT 9  
 HSA302589 1074 bp DNA PRI 06-JAN-2001  
 LOCUS Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line SA.  
 DEFINITION  
 ACCESSION AJ302589  
 VERSION AJ302589.1 GI:12054402  
 KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J.,

Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 (in) Kasahara, M. (Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)

REFERENCE  
 AUTHORS  
 Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A.,  
 Younger, R.M. and Beck, S.  
 Polymorphisms in olfactory receptor genes: a cautionary note  
 3 (bases 1 to 1074)  
 Younger, R.M., Amadou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
 Volz, A., Ziegler, A. and Beck, S.  
 Characterisation of clustered MHC-linked Olfactory Receptor Genes  
 in Human and Mouse  
 Unpublished  
 4 (bases 1 to 1074)  
 Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.  
 and Ziegler, A.  
 MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 Unpublished  
 5 (bases 1 to 1074)  
 Ehlers, A.  
 Direct Submission  
 Submitted (06-NOV-2000) Ehlers, A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY

FEATURES  
 source  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="SA"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /codon\_start=1  
 /product="olfactory receptor"  
 /protein\_id="CAC20509.1"  
 /db\_xref="GI:12054403"  
 /translation="MMWVNSVPEFLLVPSDQWLEIPPEVLEFSYLLTIFGNLT  
 IIVSHVDFKLTMYFSLNLSLDLCYTPOMLVNCRKYSVSGCYAOLF  
 IFIALGTECLLAVMCFDEFAICRPLHYSIIHMRCLPCLAAASMTSGSNVLOS  
 TWLTKMPLCGHKEVDHPECEPAIKLISCTDPTANEAELFISYLLIPVLTILISY  
 AFYQAVLRIDSAEGORAKRGTCGSHLIVSLFYGTASMLQTPSPSSKRGKAVSL  
 FCGIILAPMLNPLVLTLLRNKEVKEAFKRLVAKSLINQETIRNMOMISFAKDTVLTVTNF  
 SASCPITVTIENCNLPQRKFP"

BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN

Query Match 77.1%; Score 126.2; DB 9; Length 1074;  
 Best Local Similarity 86.3%; Pred. No. 2.8e-197;  
 Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 1 atgaattggtaaatgacagcatatcagaggttatctctggtgttccaatcga 60  
 DB 1 ATGAATTGGTAATAAGAGTGTCCACAGGAGTTCATTCTGTAGTTTTCAGATCA 60  
 QY 61 ccttgagcgaattcccaactctgtgtctcttgattcttacaactgtgacatctt 120  
 DB 61 CCATGGCTAGAGATTCACACCTTTGTGATGTTCTGTTTCTATATCTTGACATCTT 120  
 QY 121 ggaacatgcacattatcttagtgcacgcctgcgacacacacatccatccatgtat 180  
 DB 121 GGCAATCTGACAAATAATCTTGTGTACATGATTTCAAACTCCACACCCCTATGTAC 180  
 QY 161 ttttcttacaactatcatctctgtgatcttctgttacaacacatgtatccacaa 240

|||||  
 DB 161 TTTTCTTAGAATCTCTCACTCTGAGCTTTTCTTATACCAAGATGACTTCCACA 240  
 QY 241 atgcataaattatgcagaatcagaagaatcagatcagtcgtgtgacag 300  
 DB 241 ATGCTGTAAACATATGCAACACAGGAAGAAATCAGTATGAGGCTGTGAGCCAG 300  
 QY 301 ctttcatattcttgctctggggctactgaatatcttcttcggcagatgcctt 360  
 DB 301 CTTTCATTTTCTTGCCCTGGGTTCCACAGATATCTTCTGCGCTCATGTGCTTT 360  
 QY 361 gataagttgtatcattctgtgcctcccaactcactcagatcatatcaccagaagtc 420  
 DB 361 GATAGTTGTATGATTTGTTGCGCTCTCCATTTCTCATATATGATGACAGAGGCTC 420  
 QY 421 tgcctcagttgcagcgcacactcgtggatctagttttagtaactcagtggtgtct 480  
 DB 421 TGCTTCCAGTTGGCAGCTGCATCGATGATGAGCTTATGCAATTCAGTATTACAGTCC 480  
 QY 481 accctgactctcagctgcacactcgtgacacctatgtagatgacattctctgtgaa 540  
 DB 481 ACCTGACACTTAAGATGCCACTGTGTGTCACAAAGAGTGGATCATTCTTGTA 540  
 QY 541 gtcctgcactgcacagttacatctgtgtgagacacagcaacagagctgaactatcc 600  
 DB 541 GTCCCTGCTCTCTCAAGTGTCTGTGTGACACACAGCAATGAGGCTGAATTC 600  
 QY 601 ctgtcagtgagctcttccatcattatccctgcacactcattatcattatcatt 660  
 DB 601 TTCATCAGTGTCTATTCCTCTTAATACCCGACATCATCTTATATGATGCTTT 660  
 QY 661 atgtccagagcagatctgagatcagtcgtcgtgagagtcgacaaaagatctgggaa 720  
 DB 661 ATGTGCCAAGCAGTGTGGAATCCAGTCTGTGAAGTCAACGAAGCATTTGGACA 720  
 QY 721 tgtgttccatcattatgtgtgtctcttctttagtagacagcgcgtctgtgacctg 780  
 DB 721 TGTGCTCCATCTAATTTGTGTGACATTTTATGATGACATCTCCATGATACCTG 780  
 QY 781 caaccaccttcgcccagctcacaagacagaagaatgttctctctctcattatgaalc 840  
 DB 781 CACCACCTTCCAGCCAGCTCCAAAGCGGGAAGATGTTCTCTCTGTGTGAATC 840  
 QY 841 attgacacatgcgtgaatcccttatatacattaggaacagaaggttaagaagtc 900  
 DB 841 ATTGACCATCTCAATCCCTTATATATACCTTAGGACAAAGAGGTAAGAGAGCC 900  
 QY 901 ttaaaagttgtgttcg-aaagatctcttctaataagaataaa 942  
 DB 901 TTTAAAGGTTGTGCAAGAGTCTTCTTATACAGAAATAA 943

RESULT 10  
 HSA302590  
 LOCUS HSA302590 1074 bp DNA PRI 06-JAN-2001  
 DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line YAR.  
 ACCESSION AJ302590  
 VERSION AJ302590.1 GI:12054404  
 KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J.,  
 Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
 Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 (in) Kasahara, M. (Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 Springer-Verlag, Tokyo, Japan (2000)  
 2 (bases 1 to 1074)



**Fig. 1** **Genomic organization of the human olfactory receptor gene family.** The figure shows the genomic organization of the human olfactory receptor gene family. The top part of the figure shows the genomic organization of the human olfactory receptor gene family, with the gene structure and the location of the olfactory receptor gene family. The bottom part of the figure shows the genomic organization of the human olfactory receptor gene family, with the gene structure and the location of the olfactory receptor gene family.

Gene	Accession	Location	Gene Structure	Gene Name
OR1A1	U04693	1p32	10 exons, 9 introns	Olfactory receptor 1A1
OR1A2	U04694	1p32	10 exons, 9 introns	Olfactory receptor 1A2
OR1A3	U04695	1p32	10 exons, 9 introns	Olfactory receptor 1A3
OR1A4	U04696	1p32	10 exons, 9 introns	Olfactory receptor 1A4
OR1A5	U04697	1p32	10 exons, 9 introns	Olfactory receptor 1A5
OR1A6	U04698	1p32	10 exons, 9 introns	Olfactory receptor 1A6
OR1A7	U04699	1p32	10 exons, 9 introns	Olfactory receptor 1A7
OR1A8	U04700	1p32	10 exons, 9 introns	Olfactory receptor 1A8
OR1A9	U04701	1p32	10 exons, 9 introns	Olfactory receptor 1A9
OR1A10	U04702	1p32	10 exons, 9 introns	Olfactory receptor 1A10
OR1A11	U04703	1p32	10 exons, 9 introns	Olfactory receptor 1A11
OR1A12	U04704	1p32	10 exons, 9 introns	Olfactory receptor 1A12
OR1A13	U04705	1p32	10 exons, 9 introns	Olfactory receptor 1A13
OR1A14	U04706	1p32	10 exons, 9 introns	Olfactory receptor 1A14
OR1A15	U04707	1p32	10 exons, 9 introns	Olfactory receptor 1A15
OR1A16	U04708	1p32	10 exons, 9 introns	Olfactory receptor 1A16
OR1A17	U04709	1p32	10 exons, 9 introns	Olfactory receptor 1A17
OR1A18	U04710	1p32	10 exons, 9 introns	Olfactory receptor 1A18
OR1A19	U04711	1p32	10 exons, 9 introns	Olfactory receptor 1A19
OR1A20	U04712	1p32	10 exons, 9 introns	Olfactory receptor 1A20
OR1A21	U04713	1p32	10 exons, 9 introns	Olfactory receptor 1A21
OR1A22	U04714	1p32	10 exons, 9 introns	Olfactory receptor 1A22
OR1A23	U04715	1p32	10 exons, 9 introns	Olfactory receptor 1A23
OR1A24	U04716	1p32	10 exons, 9 introns	Olfactory receptor 1A24
OR1A25	U04717	1p32	10 exons, 9 introns	Olfactory receptor 1A25
OR1A26	U04718	1p32	10 exons, 9 introns	Olfactory receptor 1A26
OR1A27	U04719	1p32	10 exons, 9 introns	Olfactory receptor 1A27
OR1A28	U04720	1p32	10 exons, 9 introns	Olfactory receptor 1A28
OR1A29	U04721	1p32	10 exons, 9 introns	Olfactory receptor 1A29
OR1A30	U04722	1p32	10 exons, 9 introns	Olfactory receptor 1A30
OR1A31	U04723	1p32	10 exons, 9 introns	Olfactory receptor 1A31
OR1A32	U04724	1p32	10 exons, 9 introns	Olfactory receptor 1A32
OR1A33	U04725	1p32	10 exons, 9 introns	Olfactory receptor 1A33
OR1A34	U04726	1p32	10 exons, 9 introns	Olfactory receptor 1A34
OR1A35	U04727	1p32	10 exons, 9 introns	Olfactory receptor 1A35
OR1A36	U04728	1p32	10 exons, 9 introns	Olfactory receptor 1A36
OR1A37	U04729	1p32	10 exons, 9 introns	Olfactory receptor 1A37
OR1A38	U04730	1p32	10 exons, 9 introns	Olfactory receptor 1A38
OR1A39	U04731	1p32	10 exons, 9 introns	Olfactory receptor 1A39
OR1A40	U04732	1p32	10 exons, 9 introns	Olfactory receptor 1A40
OR1A41	U04733	1p32	10 exons, 9 introns	Olfactory receptor 1A41
OR1A42	U04734	1p32	10 exons, 9 introns	Olfactory receptor 1A42
OR1A43	U04735	1p32	10 exons, 9 introns	Olfactory receptor 1A43
OR1A44	U04736	1p32	10 exons, 9 introns	Olfactory receptor 1A44
OR1A45	U04737	1p32	10 exons, 9 introns	Olfactory receptor 1A45
OR1A46	U04738	1p32	10 exons, 9 introns	Olfactory receptor 1A46
OR1A47	U04739	1p32	10 exons, 9 introns	Olfactory receptor 1A47
OR1A48	U04740	1p32	10 exons, 9 introns	Olfactory receptor 1A48
OR1A49	U04741	1p32	10 exons, 9 introns	Olfactory receptor 1A49
OR1A50	U04742	1p32	10 exons, 9 introns	Olfactory receptor 1A50
OR1A51	U04743	1p32	10 exons, 9 introns	Olfactory receptor 1A51
OR1A52	U04744	1p32	10 exons, 9 introns	Olfactory receptor 1A52
OR1A53	U04745	1p32	10 exons, 9 introns	Olfactory receptor 1A53
OR1A54	U04746	1p32	10 exons, 9 introns	Olfactory receptor 1A54
OR1A55	U04747	1p32	10 exons, 9 introns	Olfactory receptor 1A55
OR1A56	U04748	1p32	10 exons, 9 introns	Olfactory receptor 1A56
OR1A57	U04749	1p32	10 exons, 9 introns	Olfactory receptor 1A57



TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
in Human and Mouse  
JOURNAL Unpublished  
REFERENCE 4 (bases 1 to 1074)  
AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.  
and Ziegler, A.  
TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes  
JOURNAL 5 (bases 1 to 1074)  
AUTHORS Ehlers, A.  
TITLE Direct Submission  
JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY

FEATURES  
Source Location/Qualifiers  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="OLGA"  
1..1074  
/gene="6M1-10\*01"  
1..1074  
/gene="6M1-10\*01"  
/product="Olfactory receptor"  
/protein\_id="CAC20511.1"  
/db\_xref="GI:12054407"  
/translation="MMVWVNSVDFEILFVSDPMLFPPFVWFLPSYLLTFTGNT  
IIIVSHDFELHTPMYFPLNSLIDLCYTSYVPMALVINCRTKYSYGGCAOLF  
IFLAGSTECILAVMCFDRFVACRPLHYSIIHMRCLFOLAAASMSISGNSVLDS  
TWLTKMPLCGHKEVDHFCVPLKTSKTDANERFESVLLPVLITLISY  
AFIVQAVLRIOASAGOKRAGTOSSHIIVSLPRTGISMTLOPPSSSDRGKMWEL  
FCGLIAPMLNPLVLTLEKKEAFKRLVAKSLNDEIRNMOMISFAKDVLYLTINP  
SASCPFIIVIIENYCNLPORRFP"

BASE COUNT 278 a 248 c 202 g 346 t  
ORIGIN

Query Match 77.1%; Score 726.2; DB 9; Length 1074;  
- Best Local Similarity 86.3%; Pred. No. 2.8e-197;  
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 1 atgaattggtgaatagacagcatcatagagagattatctgcgtgggtcttcagatcga 60  
|||||  
DB 1 ATGAATTGGTGAATAAAGAGTCCACAGAGATTTCATTGTTTCTTCAGATCAA 60  
QY 61 ccttgagctgaggttccacactccttggtgtcttctgattcttaccacgtgacatctt 120  
|||||  
DB 61 CCATGGCTAGAGATTCCACCCTTGTGTGATGTTTCTGTTTCTTCTATATCTTGACAACTCTT 120  
QY 121 ggcacatcgacatattatagtgacagcctgagacacaaatcatatccccatglat 180  
|||||  
DB 121 GGCAATCGACAAATATCTGTGTGACATGATGATTTCAAATCCACACCCCATATGTAC 180  
QY 181 tttttcttaccatctatcaactcctctgagatcttggttaccacacatgtaacgtccaca 240  
|||||  
DB 181 TTTTCTTACCAATCTCTCACTCTCTGACCTTGTGCTTATACCAAGATGACAGTTCCACAA 240  
QY 241 atgctagtaaatattagcagcagagaaatgaatcatcagttatcggtgctgttaaccag 300  
|||||  
DB 241 ATGCTGAGTAACATATGACACACGAAAGTAATCATCTATGATGCTGTGTGACCCAG 300  
QY 301 cttctcatattctgagcctggtgggtactgaatatcttctcctgagcgtcatgctctt 360  
|||||  
DB 301 CTTTTCATTTTCTGCGCTTGGCTTCCACAGAAATGCTTTCCTGCGCATGTCGCTT 360  
QY 361 gataggtttgagctatttgcggtcttaccatctactactcagttatcagtaacagagatc 420  
|||||  
DB 361 GATAGGTTTGAAGCTATTGTGTGGCTCTCCATTAATCAATTAATCATGACACAGAGGCTC 420  
QY 421 tgcctcagattgagcagccgcatccttggttactggttttagtaactcagtggtgtcct 480

|||||  
DB 421 TGCTTCCAGTTGGCAGCTGCATCCGATGATGAGCTTTAGCAATTCAGTATACAGTCC 480  
QY 481 accctgactctccagctgctcaactctgtaaccccatagtgaatgatacacttctctgtaa 540  
|||||  
DB 481 ACCCTGACTCTCCAGCTGCATCCGATGATGAGCTTTAGCAATTCAGTATACAGTCC 540  
QY 541 gtccctgagcagctcaaatatcttggtgtgagacaacagagaatgagagcgaactatc 600  
|||||  
DB 541 GTCCCTGCTTGTCTCAATGTTGTCTGTGTGACACACGAAATGAGGCTGACTATTC 600  
QY 601 ctgtcagtgagctcttccatctaatatccctgacacatcatatcatatcatgctttt 660  
|||||  
DB 601 TTCATCAGTGTGATATCTCTCTTATATACCCGAGACACTGATCATTATATGATGATCTTT 660  
QY 661 attgtcagagcagatattgaagatatagctctgtaaggtgtaacaaagaacttggagca 720  
|||||  
DB 661 ATTGTCCAGAGTGAAGTGAATCCAGCTCTGAGGTCACAGAAAGGCAATTTGGGACA 720  
QY 721 tgtgttccatctaatgtgtgtgtctcttctttagtagtagagcgtctgtgactg 780  
|||||  
DB 721 TGTGCTCCCATCTATATGTTGTGTGCTCACTTTTATGTTAGTACGATATCTCATGTACTG 780  
QY 781 caaccaccttgcagcagctcccaagacccaagaagaatggttctctctctatagaatc 840  
|||||  
DB 781 CAACCACCTTCACCCAGCTCCAAAGACCGGGAAAGATGGTTCTCTCTGTGGAGATC 840  
QY 841 attgacccatgctgaatcccttatatatcatcacttagaacaagaagtaagaagagc 900  
|||||  
DB 841 ATTGACCCATGCTGAATCCCTTATATATACCTTAGGAAACAAAGAGTAAAGAGACC 900  
QY 901 tttaaaaggttggtgtgc-aagagcttctttaaagaagataa 942  
|||||  
DB 901 TTTAAAGGTTGTTGCAAGAGTCTCTTATATCAAGAAATTA 943

RESULT 12  
HSA302592  
LOCUS HSA302592 1074 bp DNA PRI 06-JAN-2001  
DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line AMa1.  
ACCESSION AJ302592  
VERSION AJ302592.1 GI:12054408  
KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1074)  
AUTHORS Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J.,  
Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(in) Kasahara, M. (Ed.);  
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
FUNCTION: 110-130;  
Springer-Verlag, Tokyo, Japan (2000)  
2 (bases 1 to 1074)  
REFERENCE Ziegler, A., Ehlers, A., Forbes, S.A., Trowsdale, J., Volz, A.,  
Younger, R.M. and Beck, S.  
TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1074)  
AUTHORS Younger, R.M., Amandou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
Forbes, S.A., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J.,  
Volz, A., Ziegler, A. and Beck, S.  
TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
in Human and Mouse  
JOURNAL Unpublished  
REFERENCE 4 (bases 1 to 1074)  
AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.  
and Ziegler, A.  
TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes

JOURNAL Unpublished  
REFERENCE 5 (bases 1 to 1074)  
AUTHORS Ehlers, A.  
TITLE Direct Submission  
JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY

## FEATURES

## source

Location/Qualifiers  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="AMA1"  
1..1074  
/gene="6M1-10\*01"  
1..1074  
/gene="6M1-10\*01"  
/codon\_start=1  
/product="Olfactory receptor"  
/protein\_id="CAC20512.1"  
/db\_xref="GI:12054409"  
/translation="MNNYKSVROEFLILVSDPMLIEPFWMLFSYITITGNLT  
IILVSHDFKLTPTMYTFLSNLSLDLCYTTSTVPMLVNTCNTRKVISIGGVAQLF  
IFLALGTECELLAVMCFDFVAICRPLHYSIIMHRCFGLAASWISGFSNVLDS  
TWLKMPLGHEVDHFEVPALEKSCVDYTNAELEFISYFLILILISY  
AFIVQAVLRISAGORAKAFSTGSHLIVLVEFGTAISMLOPPSPSKRMVSL  
FCGIIAPMLNLIYTLRNKEVKAERFLVAKSLNIOEIRNNQMSPKADIVLYLTNP  
SASCPITVITENYCNLPQRKEP"

## gene

## CDS

BASE COUNT 278 a 248 c 202 g 346 t  
ORIGIN

Query Match 77.1% Score 726.2; DB 9; Length 1074;  
Best Local Similarity 86.3%; Pred. No. 2.8e-197;

Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

OY 1 atgaattgggtaaatgaacagcatcatcaggaagattatctgctgggttctcagatcga 60  
DB 1 ATGAATTGGGTAATAAGAGTGTCCACGAGATTCATCTGTAGTTTCTTCACATCA 60  
OY 61 ccttggcggagattcctcctcttctgctctctgattcttcttaacgtgtgacctt 120  
DB 61 CCATGGCTAGAGATTCCACCTCTGTGATGTTCTGTTTCTATATCTTGACAAATCTTT 120  
OY 121 ggaactcggaccattatctagtgcaagcctgcagacacacacattcatacccatgtat 180  
DB 121 GGCATCTGACAAATAATCTTGTGTACATGTGATTTCAATCTCACACCCCTATGTAC 180  
OY 181 ttttcttaccatctatctactcctgtgattctgtttacacacacatgtagtccacaa 240  
DB 181 TTTTCTTACCAATCTCACTCCTCGACCTTGTCTATACCAACAAGTACATTCACACA 240  
OY 241 atgcagtaaatattatgcaagcagcgaagaatcaagttatcgtgctgtgtgacccag 300  
DB 241 ATGCTGTAATAATATGCAACACGAGAAATACAGTTAAGTGCGCTGTGCGCCAG 300  
OY 301 ctttcatattctggccttgggggcttactgaaatacttctcctggcgtcatgctctt 360  
DB 301 CTTTTCATTTCTGCTGGCTTGGGCTTCCACAGAAATCTTCTTCTGCGGTATGCTTT 360  
OY 361 gataagtttgtagctattgttcgacctcctcatctactactcagttatcatcagcagagac 420  
DB 361 GATAGTTTGTAGCTAATTTGTTGGGCTCTCCATTTACTCAATTAATCATGCACAGAGGCTC 420  
OY 421 tgcctcagttggcagcgcgacactcctggttactctgttcttagtaactcagtggtgtct 480  
DB 421 TGCCTCCAGTTGGCAGCTGCATCTCGATGTAGTGCTTACGAATTCAGTATTACAGTCC 480  
OY 481 acccagactccagctccagctcgtgaccccatgtagatagatcatcttctcgtgaa 540  
DB 481 ACCTGGACACTTAAGATGCACCTGTGTGTGCACAAGAAGTGGATCACTTCTTGTGTAA 540  
OY 541 gtccctgactgctcaagtctctgtgtgtgagacacagcaaatgaggtcgaactatc 600

DB 541 GTCCCTGCTCTGCTCAAGTGTGCTGTGTGACACACAGCAAAATGAGGCTCACTATTC 600  
OY 601 ctgtcagtgagctcttccatctaatatccctgacacccaacccatatacatatgctt 660  
DB 601 TTCATCAGTGTGCTATCTCTTCTTAATACCCGACACATCACTTATATGATCTTTT 660  
OY 661 attgtccgagcagtatagagatagacagctgtgtaagttcgacaaagacatttggaca 720  
DB 661 ATTGTCCAAGCAGTGTGAGAAATCCAGTCTGCTGAAGTCAACGAAGACATTTGGACA 720  
OY 721 tgtgtctccatctaatctgtgtgtctcttcttcttaagtagacagcgcctgtgtgacctg 780  
DB 721 TGTGCTCCCATCTAATGTGTGTGTGCTACATTTTATGTAGACAGTATCTCCATGTACCTG 780  
OY 781 caaccacacttgcagcagctcccaaggaacgaagaatggttctctctctctatgaaac 840  
DB 781 CAACCACTTTCACCCAGCTCCAAAGACCGGGAAGATGTTTCTCTCTGTGTGAATC 840  
OY 841 attgacccatctgaatcccttatatactatagacacttaggaacagaggtgaagaagc 900  
DB 841 ATTGCACCCATGCTAATCCCTTATATATACACTTAGACAAAGAGTTAAGAGAGCC 900  
OY 901 tttaaaaggttgtgtgc-aagagctctcttaatcaagaataa 942  
DB 901 TTTAAAGGTTGCTGCAAGAGCTCTCTTATCAAGAAATTA 943

## RESULT 13

## HS193B12/C

## LOCUS

## DEFINITION

Human DNA sequence from PAC 193B12 on chromosome 6p21.3-22.3. Contains histone H2A/d, H2B/d, H2A.1, H1.5, H3.F, H4.K, H3/J genes, histone H2B.1 and hypothetical protein H4 pseudogenes, histone H2A.1 and H2B.2 duplicate genes, Glycine (GGC) tRNA gene, olfactory receptor OLI like gene, ESTs STSs and predicted CpG islands.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Direct Submission  
Submitted (12-DEC-1997) Chromosome 6 Project Group  
(http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquary@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Apr 24, 1998 this sequence version replaced g1:2578067.  
IMPORTANT: This sequence is the entire insert of clone 193B12.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variations annotated may not be found in the sequence submission  
corresponding to the overlapping clone as we submit sequences with  
only a small overlap as described above.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome 6, constructed in collaboration by the Sanger  
Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler,  
David Ruddy, Jeffrey Gruen.  
Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6/ This sequence has been finished  
according to sequence map criteria as follows. An attempt is made  
to resolve all sequencing problems, such as compressions and  
repeats, but not necessarily within known annotated human repeat  
sequence elements (e.g. Alu). Where the sequence is ambiguous,  
there is an annotation using the 'unsure' feature key.  
The true left end of clone 193B12 is at 1 in this sequence. The

true right end of clone 193B12 is at 100375.  
193B12 is from the library RPCI1 constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong.  
For further details see <http://bacpac.med.buffalo.edu/>.

## FEATURES

```

source
    1..100375
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="6"
        /map="6p21.3-22.3"
        /clone="RPI-193B12"
        /clone_id="RPCI-1"
        211..312
repeat_region
    /note="MIR2 repeat: matches 33..137 of consensus"
    534..768
repeat_region
    /note="AluJo repeat: matches 289..35 of consensus;
incomplete repeat"
    1001..1308
    /note="match: 5' EST H54606"
    complement(<1589..>2087)
mRNA
    /gene="H2A/d"
    /note="match: ESTs AA426351 AA188780 AA145892 AA048695
AA061654 H15508 W41535 AA272868 AA453105 W65215 AA231513
AA073767 AA528953 AA492640 W11262 AA060372 AA027719
AA143419 U90551: match: CDNAS L19778 L19779 X58069 X14850
D17284: match: genomic DNAs Z83739 X83549 Z83736 Z83742
X05862 Y00117 X16148 Z30940 U62674 X57138 U62669 X02218
U62673 X07763 X07758"
    /evidence="not_experimental"
    /product="histone H2A"
    /protein_id="CA01417.1"
    /db_xref="GI:3080458"
    /translation="MSGRGKGGKARAKKATRRSRAGLQPPVGVHRLKRGNYAERY
GAGAPVLAVALLEYLAIELELAGNARNKRTIIPRHLOAIRNDELKLLGKVT
IAGGVPIQAVLPLPKTESHHKAKK"
    1807..2816
misc_feature
    /note="putative Cpg island"
    <2327..>2857
    /gene="H2B/d"
    /note="H2B.1 and H2B.2 like; match: ESTs AA354038 AA217829
W49399 AA396588 AA080125 W41523 AA218213 AA423665 AA562098
AA493023 AA237851 W41584 AA255145 AA327803 AA064568
AA066051 W73546 AA493453 AA315768 AA280257 R76857 AA256738
AA2055 N99711 N31296 N31249 W71872 AA255154 N28495
AA218213 AA286933 R79451 W41523 AA286932 N41584 AA066051
W73546 AA280257 AA507431 AA571199 AA571205 AA010223 N29733
N22907 H81180 AA405358 AA354035: match: CDNAS M25487
M18045 M18046 M60756: match: genomic DNAs Z83336 Z83740
X02621 Z80779 Z80780 X57985 U62669 Z80782"
    /evidence="not_experimental"
    /product="histone H2B"
    2327..2857
gene
    /gene="H2B/d"
    2403..2783
CDS
    /note="H2B.1 and H2B.2 like; match: proteins Q00715 P10854
P33778 P02278 P02294 Q64477 CE07075 P27795 CE05165 CE07280
P10853"
    /codon_start=1
    /evidence="not_experimental"
    /product="histone H2B"
    /protein_id="CA01418.1"
    /db_xref="GI:3080459"
    /db_xref="SPTREMBL:Q99877"
    /translation="MPEPSKSPAPKGGSKKAVTKAOKKDKKKRKRKESYSVYVYK
VLKGVHPTGTSKAMGIMNSFVNDIERIAGEASRLAHYNNKRSITISREIQTAVRL

```

```

repeat_region
    2956..3243
    /note="Alusg repeat: matches 13..303 of consensus"
    4365..4499
    /note="AluJo repeat: matches 1..134 of consensus;
incomplete repeat"
    4501..4556
    /note="14 copies of 4 mer 80 & conserved"
    4786..5085
    /note="Alusg repeat: matches 299..1 of consensus"
    6646..6723
    /note="39 copies of 2 mer 82 & conserved"
    7038..7077
    /note="20 copies of 2 mer 85 & conserved"
    7408..7702
    /note="Alusx repeat: matches 3..297 of consensus"
    10347..10644
    /note="AluY repeat: matches 2..300 of consensus"
    11933..11977
    /note="MIR2 repeat: matches 100..146 of consensus"
    12051..12094
    /note="22 copies of 2 mer 98 & conserved"
    12459..12640
    /note="AluJo repeat: matches 120..299 of consensus;
incomplete repeat"
    12834..13127
    /note="Alusg repeat: matches 1..296 of consensus"
    15893..16409
    /note="match: BAC end B47865"
    17170..17463
    /note="AluY repeat: matches 293..1 of consensus"
    17841..18149
    /note="match: 5' ESTs D31045 D31402"
    18070..18215
    /note="2 copies of 73 mer 95 & conserved"
    18695..18995
    /note="Alusx repeat: matches 1..302 of consensus"
    19116..19178
    /note="MIR repeat: matches 173..110 of consensus"
    20099..20194
    /note="MIRB repeat: matches 1..95 of consensus"
    20197..20490
    /note="AluY repeat: matches 301..8 of consensus"
    20495..20572
    /note="MIRB repeat: matches 84..161 of consensus"
    20574..20881
    /note="AluYB8 repeat: matches 308..1 of consensus"
    20912..20996
    /note="FLAMC repeat: matches 1..85 of consensus"
    21030..21345
    /note="MIR1C repeat: matches 169..466 of consensus"
    22763..23053
    /note="MER2 repeat: matches 345..19 of consensus"
    23261..23657
    /note="MIR1F repeat: matches 480..44 of consensus"
    23759..24062
    /note="Alusg repeat: matches 1..303 of consensus"
    25062..25189
    /note="MIR2 repeat: matches 2..137 of consensus"
    26058..26355
    /note="Alusx repeat: matches 302..8 of consensus"
    27721..27772
    /note="2 copies of 26 mer 96 & conserved"
    complement(27803..28142)
    /gene="H2B.1"
    complement(27803..28142)
CDS
    /gene="H2B.1"
    /note="match: proteins Q08859 P06899 P10854 P23527 P33778
O64524 P06900 P02279 P02278 P02281 P10853 Q64477: match:
ESTs AA256738 AA327803 AA286933 X131568 AA286932 AA280257
AA493453 AA423665: match: CDNAS M18046 M18045 M60976
M25487: match: genomic DNAs X83547 Z83336 Z80780 U62669
X59961 U77510 X57138"

```



\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 17 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1025: contig of 1025 bp in length
1026 1125: gap of 100 bp
1126 3025: contig of 1900 bp in length
3026 3125: gap of 100 bp
3126 4346: contig of 1221 bp in length
4347 4446: gap of 100 bp
4447 7420: contig of 2974 bp in length
7421 7520: gap of 100 bp
7521 10669: contig of 3149 bp in length
10670 10769: gap of 100 bp
10770 14258: contig of 3489 bp in length
14259 14358: gap of 100 bp
14359 18958: contig of 4600 bp in length
18959 19058: gap of 100 bp
19059 23968: contig of 4910 bp in length
23969 24068: gap of 100 bp
24069 28546: contig of 4478 bp in length
28547 28646: gap of 100 bp
28647 37679: contig of 9033 bp in length
37680 37779: gap of 100 bp
37780 45607: contig of 7828 bp in length
45608 45707: gap of 100 bp
45708 54439: contig of 8732 bp in length
54440 54539: gap of 100 bp
54540 63730: contig of 9191 bp in length
63731 63830: gap of 100 bp
63831 77842: contig of 14012 bp in length
77843 77942: gap of 100 bp
77943 98659: contig of 20717 bp in length
98660 98759: gap of 100 bp
98760 127843: contig of 29084 bp in length
127844 127943: gap of 100 bp
127944 166758: contig of 38815 bp in length.
  
```

## FEATURES

```

source
1. 166758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-59801"
/clone_11p="RPC1-11 Human Male BAC"
1. 1025
/misc_feature
/note="assembly_fragment"
1126. 3025
/misc_feature
/note="assembly_fragment"
3126. 4346
/misc_feature
/note="assembly_fragment"
clone_end:77
vector_side:left"
4447. 7420
/misc_feature
/note="assembly_fragment"
7521. 10669
/misc_feature
/note="assembly_fragment"
10770. 14258
/misc_feature
/note="assembly_fragment"
14359. 18958
/misc_feature
/note="assembly_fragment"
19059. 23968
/misc_feature
/note="assembly_fragment"
24069. 28546
/misc_feature
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
28647. 37679
/misc_feature
/note="assembly_fragment"
  
```

```

misc_feature 37780..45607
/note="assembly_fragment"
misc_feature 45708..54439
/note="assembly_fragment"
misc_feature 54540..63730
/note="assembly_fragment"
misc_feature 63831..77842
/note="assembly_fragment"
misc_feature 77943..98659
/note="assembly_fragment"
misc_feature 98760..127843
/note="assembly_fragment"
misc_feature 127944..166758
/note="assembly_fragment"
BASE COUNT 48778 a 33823 c 33960 g 48591 t 1606 others
ORIGIN
Query Match 77.1%; Score 726.2; DB 2; Length 166758;
Best Local Similarity 86.3%; Pred. No. 4e-197;
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;
QY 1 atgaattgggaataagacagatcatatagagagattatctcgtggtttctcagatga 60
|||||
Pb 99340 ATGAATTGGTAATAAGAGAGTCCACAGAGTTCATTCGTGTTCTGCAGATCAA 99399
QY 61 ccttgagcgtgagttccacccctctggtgtctcttgatttcttaacctgacacctt 120
|||||
Pb 99400 CCATGGCTAGAGATGCCACCCCTTGTGATGTTTCCTTATATCTTGACAAATCTTT 99459
QY 121 ggaactcgaacatattctagttcagcgtcgtgacacacacatcataccacagat 180
|||||
Pb 99460 GGCAATCTGACAAATATCTTGTGTGCACATGTGATTTTCAAACTCACACCCCTATGTAC 99519
QY 181 ttttcttacaactatcatcctcctgtagcttctgttacaacacatgacagttccaca 240
|||||
Pb 99520 TTTTCTTACCAATCTCTCACTCCTGACCTTCTCTATACCAAGATGTTCCACAA 99579
QY 241 atgcagtaaatattatgcagatcatcagaaagtaaatcatatcgtggtgtgtgaccag 300
|||||
Pb 99580 ATGCTGATTAACATATGTAACACCGAGAAAGTAATCACTTATAGTGTGTGTGCCAG 99639
QY 301 ctttcatattctggtccttggtggtgactgaatattcctccctggtccatgtccctt 360
|||||
Pb 99640 CTTTTCATTTTCCTGCTTGCTTGCTTCCACAGAAATCTTTCCTGCGCTCATGTGCTTT 99699
QY 361 gatagttttagtctatttgcgtcctccatcatcctcagttatcatatgacacagagactc 420
|||||
Pb 99700 GATAGTTTGTAGTATTTGTGCGCTCCTCCATTCATTAATATCATGACACAGAGGCTC 99759
QY 421 tgcctcagttggcagccgcatccctggttactggttttagtaactcagtggtgtct 480
|||||
Pb 99760 TGCCTCCAGTTGGCAGCTGCATCTGGATTTAGTGGCTTTAGCAATTCAGTATACAGTCC 99819
QY 481 accctgactccagctgcacactctgtgaccccttgatagatgacacttctcgtgaa 540
|||||
Pb 99820 ACCCTGACACTTAAGATGCCACTGTGTGTACCAAGAGATGATCACTTCTTGTGAA 99879
QY 541 gtcctgacagttcagattatcttctgtgtgagacaacagaaatgaggtcgtgacattc 600
|||||
Pb 99880 GTCCTGCTGCTGCTCAAGTTGCTGTGTGACACACAGCAAGAGAGGCTGAATATTC 99939
QY 601 ctgtcagttggtcttccatctaatatccctggaacatccatctatataatgctttt 660
|||||
Pb 99940 TTCATCAGTGTCTATTCCTCTTAATACCCGACACATCCTTAATATCGATGCTTTT 99999
QY 661 attgctcgagcagttatgagatagacagctcgtgaggtcgcgacaaagacatttgagca 720
|||||
Pb 100000 ATTGTCCAAGCAGCTGTGAGATTCAGATCTGTGTAAGGTCAACGAAGCATTTGGGACA 100059
QY 721 tgtgttccatctaatgtgtgtctctttttagtagtaacagcgtctcgtgtactctg 780
|||||
Pb 100060 TGTGCTCCCATCTAATGTGTGTACCTTTTATGTGTACAGATCTCCATGTACCTG 100119
  
```

```

QY      781 caaccacttcgcccagctcccaaggaagaagatgtctctctctcatgaatc 840
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100120 CAACGACCTTCAACCCAGGCTCAAGACGGGAGAAAGATGTTCTCTCTGTGGAATC 100179
QY      841 attgcacacatcgtgaatcccttatatatacacttaggaacagaaggtaaagaagac 900
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100180 ATTGCACCCCTCTGATCCCTTATATATACACTTAGGACAAAGAGCTTAAGGAAAGCC 100239
QY      901 tttaaaaggtgtgtgc-aagagctctcttaataagaataa 942
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100240 TTAAAAAGGTGTGTCACAAAGATCTTATATACAGAAATAA 100282

RESULT 15
AC025941
LOCUS      AC025941 176277 bp DNA HTG 26-MAY-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-635011 map 6, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACCESSION  AC025941
VERSION     AC025941.2 GI:7408057
KEYWORDS    HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 176277)
AUTHORS     Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE       Homo sapiens chromosome 6, clone RP11-635011
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 176277)
AUTHORS     Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
             Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
             Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
             Campotiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
             Collymore, A., Cooke, P., Deavrellano, K., Dewar, K., Diaz, J. S.,
             Dodge, S., Domano, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
             Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
             Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
             Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kamp, L., Karatas, A.,
             Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
             Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
             McCarthy, M., McEwan, P., McGuirk, A., McKenna, K., McPheters, R.,
             Medlir, J., Meneus, L., Mihova, T., Mirand, C., Mlenga, V., Morrow, J.,
             Murphy, T., Naylor, T. J., Norman, C. H., O'Connor, T., O'Donnell, P.,
             O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
             Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
             Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
             Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
             Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
             Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
             Young, G., Zainoun, J., Zimmer, A. and Zody, M.
             Direct Submission
             Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
             On Apr 4, 2000 this sequence version replaced gi:7259782.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L8577
Center clone name: 635-O_11

----- Summary Statistics
Sequencing vector: M13: M77815: 100% of reads
Assembly program: Phrap: version 0.960731
Consensus quality: 168929 bases at least Q40
Consensus quality: 172827 bases at least Q30

```

```

Consensus quality: 174151 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 174877; sum-of-contigs
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2997: contig of 2997 bp in length
2998 3097: gap of 100 bp
3098 5606: contig of 2509 bp in length
5607 5706: gap of 100 bp
5707 9209: contig of 3503 bp in length
9210 9309: gap of 100 bp
9310 13614: contig of 4305 bp in length
13615 13714: gap of 100 bp
13715 20639: contig of 6925 bp in length
20640 20739: gap of 100 bp
20740 26545: contig of 5806 bp in length
26546 26645: gap of 100 bp
26646 37068: contig of 10423 bp in length
37069 37168: gap of 100 bp
37169 48083: contig of 10915 bp in length
48084 48183: gap of 100 bp
48184 59015: contig of 10832 bp in length
59016 59115: gap of 100 bp
59116 70420: contig of 11305 bp in length
70421 70520: gap of 100 bp
70521 80466: contig of 9946 bp in length
80467 80566: gap of 100 bp
80567 92325: contig of 11759 bp in length
92326 92425: gap of 100 bp
92426 110572: contig of 18147 bp in length
110573 110672: gap of 100 bp
110673 131898: contig of 21226 bp in length
131899 131998: gap of 100 bp
131999 176277: contig of 44279 bp in length.

FEATURES
Source
1. 176277
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone_id="RP11-635011"
/clone_idb="RPC1-11 Human Male BAC"
1. 2997
/note="assembly_fragment"
3098 . 5606
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
5707 . 9209
/note="assembly_fragment"
9310 . 13614
/note="assembly_fragment"
13715 . 20639
/note="assembly_fragment"
20740 . 26545
/note="assembly_fragment"
clone_end:n7
vector_side:right"
26646 . 37068
/note="assembly_fragment"
37169 . 48083
/note="assembly_fragment"
48184 . 59015
/note="assembly_fragment"

```

```

misc_feature      59116..70420
                   /note="assembly_fragment"
misc_feature      70521..80466
                   /note="assembly_fragment"
misc_feature      80567..92325
                   /note="assembly_fragment"
misc_feature      92426..110572
                   /note="assembly_fragment"
misc_feature      110673..131898
                   /note="assembly_fragment"
misc_feature      131999..176277
                   /note="assembly_fragment"

BASE COUNT      54493 a 33357 c 33254 g 53772 t 1401 others
ORIGIN

Query Match      77.18; Score 726.2; DB 2; Length 176277;
Best Local Similarity 86.38; Pred. No. 4e-197;
Matches 814; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 1 atgaattggtgaatagacagcatcatatagagagattatctgctgggttctcagatcga 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41072 ATGAATTTGGTTAAATGAAGTGTCCACAGAGGATTCGTGATTTTCTCAGATCAA 41131

QY 61 ccttgctgaggttccagctccttggtgtctcttctgattcttctacacgtgacatctt 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41132 CCATGGCTAGAGATTCCACCCCTTGTGATGTTCTGTTTCTTATATCTTGACATCTTT 41191

QY 121 ggcacatctgacatattatctagtgacgcctggacacacacatccatccatglat 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41192 GGCATCTGACAAATTAATCTTGTCACATGATGATTTCAAACTCCACACCCCTATGTAC 41251

QY 181 tttttcttaccatctctacactcctctgatacttctgttaccacacatgtacagtcacaa 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41252 TTTTCTTACCAATCTCTCACTCCTGGACCTTGTCTATACCAAGTACAGTTCCACAA 41311

QY 241 atgctagaataattatgcagcatcaggaagaatgaatcagttatcgtgctgtgacccag 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41312 ATGCTGGTAAACATATGCAACACGAAAGATACATATATGCTGTGCTGTGSCCAG 41371

QY 301 ctttcatattcttgaccttgaggcttggaatactctctcctggccgtcatgtcctt 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41372 CTTTTCATTTTCCTGCTGCTGGCTTCACAGAAATGTCCTTCCTGCGCTCATGTGCTTT 41431

QY 361 gataggtttagatctatttctgcgcctctccatctacgttatcatcagaccagaagctc 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41432 GATAGGTTTGAAGTATTTGCGGCTCTCATTAATCATATATCATCACAGAGGCTC 41491

QY 421 tggccctcagttggcagccgcatccctgggttactggttttagtaactcagtggtgct 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41492 TGCTTCCAGTTGGCAGCTGCATCTCGAATTAAGTGGCTTTAGCAATTCAGTATACAGTCC 41551

QY 481 accctgacttccagctgacactctgtgacccctatgtatagatagacttctctgtgaa 540
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41552 ACCGTGACACTTAAGATGCCACTGTGTGTGCACAAAGAGTGATCACTTCTTGTA 41611

QY 541 gtccctgacgtcgaagttatctgtgtgttgagacacagcaaatgaggtcgaactatc 600
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41612 GTCCCTGCTCTGCTCAAGTTGCTGTGTGTGACACACAGCAAAATGAGGCTGAACATAATTC 41671

QY 601 ctgtcagtgagctctctcatctaatatccctgagacatccatccatataatgctttt 660
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41672 TTTCAATCAATGTGCTATTTCTTCTTATATACCGTGACACTCATCTTATATCGATGCTTTT 41731

QY 661 attgctgagcagttatgagagatacagtcgtcgaagtcgacaaagaacattgggaca 720
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41732 ATTGTCGAAGCAGGTGTGAGATCCAGTCTGTGAAGTCAAAGCAAGGCAATTTGGACA 41791

QY 721 tgtggttcccatctaatgtgtgtgtctctttttatagtaacagccgtctgtgtactctg 780
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41792 TGTGCTCCCATTAATATGTGTGTGACCTTTTATGTGTACAGCTATCTCCATGTACTGTG 41851

QY 781 caaccaccttgcgccagctccaaggaagcaagaagatggttctctctctatgtaatc 840

```

```

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41852 CAACACCTTACCCAGCTCCAAAGACCGGGCAAGATGTTTCTCTCTGTGTGAATC 41911

QY 841 attgacccatgctgtaatcccttatatatatacacttaagaaagaggtlaaaggaagc 900
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41912 ATTGCAACCATGCTGAATCCCTTATATATACACTTAGAACAAAGAGGTAAGGAAGCC 41971

QY 901 tttaaaggttggtgctc-aagatctcttctaacaagaatga 942
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41972 TTTAAAGGTTGTTGCAAGAGTCTTCTTATCAAGAAATGA 42014

```

Search completed: February 26, 2002, 05:04:54  
 Job time: 5022 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 04:13:22 : Search time 114.05 Seconds

(Without alignments)  
7081.108 Million cell updates/sec

Title: US-09-755-017-1

Sequence: 1 atgaattggtaaatgacag.....tcttcttaacgaagaataa 942

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N\_Geneseq\_1101:\*

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	942	100.0	942	22	AA508541
2	942	100.0	1488	22	AA508542
3	937.4	99.5	939	22	AAH32018
4	726.2	77.1	1071	22	AAH31669
5	726.2	77.1	1071	22	AAH32250
6	641.6	68.1	648	22	AAH32231
7	537.2	57.0	1442	21	AACT7475
8	534.2	56.7	939	22	AAH32304
9	400.4	42.5	936	22	AAH31645
10	370	39.3	1002	22	AAH32063
11	364.2	38.7	1011	22	AAH32306

12	356.6	37.9	933	22	AAH32308	Human olfactory re
13	356	37.8	936	22	AAH32314	Human olfactory re
14	354	37.6	927	22	AAH32067	Human olfactory re
15	352.4	37.4	936	22	AAH32045	Human olfactory re
16	350.2	37.2	1014	21	AAH32316	Human olfactory re
17	350.2	37.2	1440	21	AACT6960	Human ORFX ORF2515
18	341.6	36.3	936	22	AAH31632	Human olfactory re
19	341.2	36.2	979	22	AAH32055	Human olfactory re
20	335.8	35.6	909	22	AAH32043	Human olfactory re
21	332.6	35.3	1629	22	AAH08960	Human novel 32164
22	328	34.8	1002	22	AAH32238	Human olfactory re
23	327.2	34.7	948	22	AAH31841	Human olfactory re
24	327.2	34.7	948	22	AAH32375	Human olfactory re
25	322.4	34.2	948	22	AAH32188	Human olfactory re
26	318	33.8	954	22	AAH31588	Human olfactory re
27	318	33.8	957	22	AAH32072	Human olfactory re
28	318	33.8	957	22	AAH32088	Human olfactory re
29	315.6	33.5	1450	21	AACT7005	Human ORFX ORF2560
30	314.4	33.4	1087	22	AAH32312	Human olfactory re
31	313.6	33.3	817	22	AAH32390	Human olfactory re
32	311	33.0	933	22	AAH31783	Human olfactory re
33	308.6	32.8	960	22	AAH32023	Human olfactory re
34	306.2	32.5	971	22	AAH32318	Human olfactory re
35	300.8	31.9	1014	22	AAH31672	Human olfactory re
36	294.6	31.3	948	22	AAH32437	Human olfactory re
37	293.6	31.2	923	22	AAH31878	Human olfactory re
38	291.2	30.9	997	22	AAH32019	Human olfactory re
39	289.8	30.8	810	22	AAH32174	Human olfactory re
40	283	30.0	921	22	AAH35759	Human partial olfa
41	282.2	30.0	951	22	AAH31879	Human olfactory re
42	280.6	29.8	930	22	AAH35755	Human olfactory re
43	280.6	29.8	951	22	AAH31584	Human olfactory re
44	278.2	29.5	1030	22	AAH31633	Human olfactory re
45	277.2	29.4	1068	22	AAH31630	Human olfactory re

#### ALIGNMENTS

RESULT 1

AA508541

ID AA508541 standard; cDNA; 942 BP.

XX

AC AA508541;

XX

XX

XX 26-SEP-2001 (first entry)

XX

XX

DE DNA encoding novel human G-protein coupled receptor (NGPCR).

XX

XX Human: novel G-protein coupled receptor; NGPCR; drug discovery;

KW diagnostic; ss.

KW

XX

XX Homo sapiens.

OS

XX	Key	Location/Qualifiers
XX	CDS	1..942
XX	FT	/\*tag= a
XX	FT	/product= "Novel G-protein coupled receptor (NGPCR)"
XX	XX	MO200151634-A1.
XX	PD	19-JUL-2001.
XX	XX	05-JAN-2001; 2001WO-US00589.
XX	PF	12-JAN-2000; 2000US-015764.
XX	PR	(LEXI-) LEXICON GENETICS INC.
XX	PA	Walke DW, Wilganowski NL, Turner CA, Friedrich G, Abulin A;
XX	PI	Zambrowicz B, Sands AT;
XX	XX	WPI: 2001-442145/47.



```

OY 1 atgaattggtgaatgacagcatcatagaggaattatctctgctggttctcagatcga 60
DB 198 atgattggtgtaaatgacagcatcatagaggaattatctctgctggttctcagatcga 257
OY 61 ccttgctggaattcgaatcctctggttctcttgattcttacaatgacatctt 120
DB 258 ccttgctggaattcgaatcctctggttctcttgattcttacaatgacatctt 317
OY 121 ggcacatcgaccattatctagtgtcaagcctgtagacccaacttacaacccatgta 180
DB 318 ggcacatcgaccattatctagtgtcaagcctgtagacccaacttacaacccatgta 377
OY 181 tttttcttaccatctatcactcctgtagatctgttaccacacatgtagctccaca 240
DB 378 tttttcttaccatctatcactcctgtagatctgttaccacacatgtagctccaca 437
OY 241 atgctagtaaatattatgacagcatcagaaagtaacagttatcgtgctgtgtagccag 300
DB 438 atgctagtaaatattatgacagcatcagaaagtaacagttatcgtgctgtgtagccag 497
OY 301 ctttcaatattctgctctggtggtgctacgtatatactctcctggtcgtatgctctt 360
DB 498 ctttcaatattctgctctggtggtgctacgtatatactctcctggtcgtatgctctt 557
OY 361 gataggtttagctaatctgctggtcctcctcattacattacattacacagagagctc 420
DB 558 gataggtttagctaatctgctggtcctcctcattacattacattacacagagagctc 617
OY 421 tgcctcagattgtagcagccgcatcctggtttagttagttagtaactcgtgtgct 480
DB 618 tgcctcagattgtagcagccgcatcctggtttagttagttagtaactcgtgtgct 677
OY 481 accctgactcctcagctgctcactctgtgacccctatgtgtagatcacttctcgtgaa 540
DB 678 accctgactcctcagctgctcactctgtgacccctatgtgtagatcacttctcgtgaa 737
OY 541 gtccctgactcctcagctgctcactctgtgtagacacagcaaatgagctgaactatc 600
DB 738 gtccctgactcctcagctgctcactctgtgtagacacagcaaatgagctgaactatc 797
OY 601 ctgtgcagtgagctcttccatctaatatccctgtagacatcactatataatgctt 660
DB 798 ctgtgcagtgagctcttccatctaatatccctgtagacatcactatataatgctt 857
OY 661 attgtcagagcatatgtagatatacagctcgtgagagtcgacaaaagcattgggaca 720
DB 858 attgtcagagcatatgtagatatacagctcgtgagagtcgacaaaagcattgggaca 917
OY 721 tgtgttcccatctaatgtgtgtctcttttttagtagtagcgcgtctctgtgactg 780
DB 918 tgtgttcccatctaatgtgtgtctcttttttagtagtagcgcgtctctgtgactg 977
OY 781 caaccacctcgccagctcgaagcaagcaagaaagatggttctctcttcatgaaac 840
DB 978 caaccacctcgccagctcgaagcaagcaagaaagatggttctctcttcatgaaac 1037
OY 841 attgacacacagctgaaatcccttatatacactttagaagaagaagtaagaagc 900
DB 1038 attgacacacagctgaaatcccttatatacactttagaagaagaagtaagaagc 1097
OY 901 tttaaaagttggttgcagaagcttcttcaatcaagaataa 942
DB 1098 tttaaaagttggttgcagaagcttcttcaatcaagaataa 1139

```

```

RESULT 3
AAH32018
ID AAH32018 standard; DNA; 939 BP.
AC AAH32018;
XX
DT 30-JUL-2001 (first entry)
XX

```

```

DE Human olfactory receptor polynucleotide, SEQ ID NO: 591.
XX
KW Human; olfactory receptor; OR; primary scent determination;
KW Secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation; ds.
XX
OS Homo sapiens.
XX
RN WO200127158-A2.
XX
PD 19-APR-2001.
XX
PF 06-OCT-2000; 2000WO-US27582.
XX
PR 08-OCT-1999; 29US-0158615.
XX
PR 24-FEB-2000; 2000US-0184809.
XX
PA (DIGI-) DIGISCENTS.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
XX
BT Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX
DR WPI; 2001-290713/30.
XX
PT New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX
PS Claim 8; Page 417; 1857pp; English.
XX
CC The present sequence is one of a number of isolated polynucleotides
XX which encode polypeptides involved in olfactory sensation. The
XX polynucleotides can be used in screening for olfactory agonists and
XX antagonists. The methods allow for the determination of primary
XX scents and the identification of the odour receptors used to detect
XX these primary scents. The methods also enable determination of
XX secondary scents and the identification of combinations of odour
XX receptors that are involved in detecting such secondary scents.
XX This enables the construction of a scent representation (also called
XX a scent fingerprint or scent profile), which may be used to re-create
XX and edit scents. Libraries of olfactory receptors are useful for
XX determining the interaction pattern of a composition with the receptors,
XX and can be used for determining differences in the olfactory faculties
XX of different individuals.
XX
SQ Sequence 939 BP; 220 A; 227 C; 187 G; 305 T; 0 other:
XX
Query Match 99.5%; Score 937.4; DB 22; Length 939;
Best Local Similarity 99.9%; Pred. No. 3,5e-278;
Matches 938; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 atgaattggtgaatgacagcatcatagaggaattatctctgctggttctcagatcga 60
DB 1 atgaattggtgaatgacagcatcatagaggaattatctctgctggttctcagatcga 60
OY 61 ccttgctggaattcgaatcctctggttctcttgattcttacaatgacatctt 120
DB 61 ccttgctggaattcgaatcctctggttctcttgattcttacaatgacatctt 120
OY 121 ggcacatcgaccattatctagtgtcaagcctgtagacccaacttacaacccatgta 180
DB 121 ggcacatcgaccattatctagtgtcaagcctgtagacccaacttacaacccatgta 180
OY 181 tttttcttaccatctatcactcctgtagatctgttaccacacatgtagctccaca 240
DB 181 tttttcttaccatctatcactcctgtagatctgttaccacacatgtagctccaca 240
OY 241 atgctagtaaatattatgacagcatcagaaagtaacagttatcgtgctgtgtagccag 300
DB 241 atgctagtaaatattatgacagcatcagaaagtaacagttatcgtgctgtgtagccag 300
OY 301 ctttcaatattctgctctggtggtgctacgtatatactctcctggtcgtatgctctt 360

```

DR WPI: 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory

PT sensation for identifying olfactory agonists and antagonists -

XX

PS Claim 8; Page 273-274; 1857pp; English.

XX

CC The present sequence is one of a number of isolated polynucleotides

CC which encode polypeptides involved in olfactory sensation. The

CC polynucleotides can be used in screening for olfactory agonists and

CC antagonists. The methods allow for the determination of primary

CC scents and the identification of the odour receptors used to detect

CC these primary scents. The methods also enable determination of

CC secondary scents and the identification of combinations of odour

CC receptors that are involved in detecting such secondary scents.

CC This enables the construction of a scent representation (also called

CC a scent fingerprint or scent profile), which may be used to re-create

CC and edit scents. Libraries of olfactory receptors are useful for

CC determining the interaction pattern of a composition with the receptors,

CC and can be used for determining differences in the olfactory faculties

CC of different individuals.

CC

XX

SQ Sequence 1071 BP; 277 A; 248 C; 201 G; 345 T; 0 other;

Query Match 77.1%; Score 726.2; DB 22; Length 1071;  
Best Local Similarity 86.3%; Pred. No. 3.4e-213;  
Matches 814; Conservative % 0; Mismatches 128; Indels 1; Gaps 1;

Oy 1 atgaattggtgtaaatagacagcatatcacaggagtattctcgtcggtttccagatoga 60  
Db 1 atgaattggtaataaagaagtgctcccaaggagtcattctgttagtllcttcagatcaa 60  
Oy 61 ccttgcggagggtttccacctcctgttgttcttcttgatttttacaactgyaccatttt 120  
Db 61 caatggcagagatctccaccttvgtgatgtlctgttlcttatatcttgacaattctt 120  
Oy 121 ggcaactctgacatattcttagtltgacagcttggaaccaaacattacatcccattat 180  
Db 121 ggcactctgacaatatctctgtgtgacaatgttgattccaactccaaacccctatgac 180  
Oy 181 tttttcttaccaalcatcacctcctgtgatctctgtttaaccacatgtagaccacaaa 240  
Db 181 tttttctttagaactctctcacctcctgtgaccttgtctataccagaatgactgccaaa 240  
Oy 241 atgcagtaaatattatgtacagcatcaagaaatgaatcaatctgtgtgtgtatgccag 300  
Db 241 atgcgtgtaaacatatgtacaacaggaaglaaalcaattatggtgtgtgtgtgccag 300  
Oy 301 ctcttcataattcttgaccttgggagctacgttaaatctctcccgagccatgtccctt 360  
Db 301 ctcttcataattcttgaccttgggagctacgttaaatctctcccgagccatgtgtctt 360  
Oy 361 gataggtttgtagctatttgcgtgcctctccatctactacatcaatlatcatgaccagagctc 420  
Db 361 gataggtttgtagctatttgcgtgcctctccatctactacatcaatlatcatgaccagagctc 420  
Oy 421 tgcctccagttggcagccgatcctgtgtatcvtggttttagtaactaagvtgtgtct 480  
Db 421 tgcctccagttggcagctgtcacctctgtgtattgtgtctttagtaactaagvtttcc 480  
Oy 481 acccgtagctccagagctgcacactctgtgacccccatgtgataagataactctcttgaa 540  
Db 481 accgtggaactcaagaatgcccacgtgtgtgtacaaagaagaagtggtacactctcttgaa 540  
Oy 541 gtccctgcagctcaagaattatctgtgtgtgtgaacaacagcaaatgaggtcgaactatc 600  
Db 541 gtccctgtctgtcccaagtgtctgtgtgtgacacaagaagaatgaggtcgaactatc 600  
Oy 601 ctgttcagttgagctctccatctaatataacctgcgaactcatccttatcataatgcttt 660  
Db 601 ttcatcaatgtgcatatctctctaataaccctgagcaactcatccttatataatgtagcttt 660



ID	AAH32231	standard: DNA; 648 BP.
XX	AAH32231	
AC	AAH32231;	
XX		
DT	30-JUL-2001 (first entry)	
XX		
DE	Human olfactory receptor polynucleotide, SEQ ID NO: 804.	
XX		
KW	Human; olfactory receptor; OR; primary scent determination;	
KW	secondary scent determination; polypeptide library; odour receptor;	
XX	scent profile; scent fingerprint; scent representation; ds.	
OS	Homo sapiens.	
XX		
PN	WO200127158-A2.	
PD		
XX	19-APR-2001.	
PE	06-OCT-2000; 2000WO-US27582.	
XX		
PR	08-OCT-1999; 99US-0158615.	
PR	24-FEB-2000; 2000US-0184809.	
XX		
PA	(DIGI-) DIGISENTS.	
XX	(YEDA ) YEDA RES & DEV CO LTD.	
PI	Wellenson J, Smith D, Lancet D, Giusman G, Fuchs T, Yanai I;	
DR	WPI: 2001-290713/30.	
PT	New polynucleotides which encode polypeptides involved in olfactory	
XX	sensation for identifying olfactory agonists and antagonists -	
PS	Claim 8; Page 502; 1857pp; English.	
XX		
CC	The present sequence is one of a number of isolated polynucleotides	
CC	which encode polypeptides involved in olfactory sensation. The	
CC	polynucleotides can be used in screening for olfactory agonists and	
CC	antagonists. The methods allow for the determination of primary	
CC	scents and the identification of the odour receptors used to detect	
CC	these primary scents. The methods also enable determination of	
CC	secondary scents and the identification of combinations of odour	
CC	receptors that are involved in detecting such secondary scents.	
CC	This enables the construction of a scent representation (also called	
CC	a scent fingerprint or scent profile), which may be used to re-create	
CC	and edit scents. Libraries of olfactory receptors are useful for	
CC	determining the interaction pattern of a composition with the receptors,	
CC	and can be used for determining differences in the olfactory faculties	
CC	of different individuals.	
XX		
XX		
SQ	Sequence 648 BP; 144 A; 166 C; 132 G; 206 T; 0 other;	
XX		
Query Match	68.1%; Score 641.6; DB 22; Length 648;	
Best Local Similarity	99.4%; Pred. No. 2.8e-187;	
Matches 644; Conservative	0; Mismatches 4; Indels 0; Gaps	0;
OY	202 cccccgagatcttgttaaccacatgtaagcccaaaatgtagtaattatgcagc 261	
Db	1 cccccgagatcttgttaaccacatgtaagcccaaaatgtagtaattatgcagc 60	
OY	262 atcaggaagaatagatcgtatcgttgctgtgtagccagcttcatatcttgcccttg 321	
Db	61 atcaggaagaatagatcgtatcgttgctgtgtagccagcttcatatcttgcccttg 120	
OY	322 ggggctactgaatcctctcctgagccgcatgcttccttgataggttctgtagcatattgt 381	
Db	121 ggggctactgaatcctctcctgagccgcatgcttccttgataggttctgtagcatattgt 180	
OY	382 cggctctccatattcgttatcgtatcgtgacccagaaactctgcttcagttgcaagcgga 441	
Db	181 cggctctccatattcgttatcgtatcgtgacccagaaactctgcttcagttgcaagcgga 240	

QY	442	tcttgaggtactggttttagtaaacctagtggtgtgtcttaccgcgactctccagctgcga	501
Db	241	tccaaaggtctacgtgttttaagtaaccagtcgtgtgtctccaccgcgactctccagctgcga	300
QY	502	ctctctgacccctatgltatagatcaacttctctctgtgaagtccctgcactgctcaagtta	561
Db	301	ctctgtgagccctatgltatagaccacttctctctgtgaagtccctgcactgctcaagtta	360
QY	562	tctgtgtgttgagacaacagcaatgtaggctgaactatctctgtctcagttagcttccat	621
Db	361	tctgtgtgttgagacaacagcaatgtaggctgaactatctctgtctcagttagcttccat	420
QY	622	cttaaaccccccgaacatcatctctatatacatatgcttttatctgtgcgaagaatttagg	681
Db	421	cttaaaccccccgaacatcatctctatatacatatgcttttatctgtgcgaagaatttagg	480
QY	682	atacagctctgtcgtgaagtcgacaaagaacatttggagcatgtgtgtccatctaatgtg	741
Db	481	atacagctctgtcgtgaagtcgacaaagaacatttggagcatgtgtgtccatctaatgtg	540
QY	742	gtgtctctcttllatagtagacagcggtctctgtgtactgtgaacctgcaaccacttgccagctc	801
Db	541	gtgtctctcttllatagtagacagcggtctctgtgtactgtgaacctgcaaccacttgccagctc	600
QY	802	aagagccaagaaagaatggtttctctctctctcatgtgaatcatgtgacc	849
Db	601	aagagccaagaaagaatggtttctctctctctcatgtgaatcatgtgacc	648
RESULT 7			
ID	AAC77475	AAC77475 standard; CDNA; 1442 BP.	
AC	AAC77475:		
DT	08-FEB-2001	(first entry)	
XX	XX	Human OREF ORF3030 polynucleotide sequence SEQ ID NO:6059.	
DE	XX		
XX	XX	Human; open reading frame; ORFX; detection; cytosstatic; hepatotropic;	
KW	KW	vulnary; antipsoiatric; antiparkinsonian; noctropic; neuroprotective;	
KW	KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;	
KW	KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;	
KW	KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;	
KW	KW	antiviral; antibacterial; antifungal; anlrheumatic; antithyroid;	
KW	KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;	
KW	KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	
KW	KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	
KW	KW	cholesterol ester storage; systemic lupus erythematosus; infection;	
KW	KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	
KW	KW	allergy; aplastic anaemia; nocturnal hemoglobinuria; burn; wound;	
KW	KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;	
XX	XX	thrombosis; contraceptive; ss.	
OS	XX	Homo sapiens.	
XX	XX	WO200058473-A2.	
XX	XX	05-OCT-2000.	
PD	XX		
XX	XX	31-MAR-2000; 2000WO-US08621.	
XX	XX		
PR	PR	31-MAR-1999; 99US-0127607.	
PR	PR	02-APR-1999; 99US-0127636.	
PR	PR	05-APR-1999; 99US-0127728.	
XX	XX	30-MAR-2000; 2000US-0540763.	
XX	XX	(CURA-) CURAGEN CORP.	
XX	XX		
XX	XX	Shimkets RA, Teach M;	
XX	XX	WPI; 2000-602362/57.	
XX	XX		





CC receptors that are involved in detecting such secondary scents.  
CC This enables the construction of a scent representation (also called  
CC a scent fingerprint or scent profile), which may be used to re-create  
CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.

50 Sequence 939 BP; 232 A; 228 C; 185 G; 294 T; 0 other,

Query Match	56.7%	Score 534.2	DB 22	Length 939
Best Local Similarity	73.1%	Pred No. 3.6e-154		
Matches 666; Conservative	0	Mismatches 253;	Indels 0;	Gaps 0

QY	1	atgaatttggcaaaatgaaagaacatatacaagaagcttcttcgctgggttcctcaagaca	60
Db	1	atgaatttggcaaaatgaaagaagctctcccaaaaagatttatactactcttgctctcaagag	60
QY	61	ccttgagctggaggttccacatccctctggtgtctctcttgattcttaacatgtagacattt	120
Db	61	gcttggtctacaatagccccctttctgtgtccctgtttaatatatacaatacaatacatatt	120
QY	121	ggcaatctgacacatattctctagtgtaacgcctgtgaacccaactctataccccaatgat	180
Db	121	ggcaatgtgtccatcatgatgtgtgtgcatcttctgatacccaactctatactcccaatgat	180
QY	181	ttttctcttacaacatctatacaatccctcgagatctctgtttacacacaaatgtaagttcccaa	240
Db	181	ttctctctcaataatctctcatctatgactctctgctataccaacaatacagttccctcat	240
QY	241	atgcagtagaaattatgatgacatcagaagaagttaactgaatctgtgctgtgtgaagccag	300
Db	241	atgtttgtgtaaaatattgtgtgtaacaaaaagacatcagctatgtctgtgtgtgtgtgtgtgt	300
QY	301	ctttacatatcttcggcctctgggtgggtactgtaatactctctctgtgcgttaatgtctct	360
Db	301	ctctacatctcttcctggccctctgagtgactaagagtgctctctctctgtgtatgtctct	360
QY	361	gatagagtttgtagctattttgtgcgcctctccatctacatccagttatcatcagaacagagctc	420
Db	361	gacagatatgtgctgtgttttgagacccctccactatgtaatgtaacacagaattatgtgttc	420
QY	421	tgctctcaagttgtagacgccgcatctctggtgttaactgtgttttaagtaacagtggtgtgtct	480
Db	421	tgccctaaagatgtagagcctctctctcatgtctatgtttctgcgaactcagtgctgcagctc	480
QY	481	acctggaattctccagccttgccacatcttggaaccccatgtgtaagaatgaactcttctgtgaa	540
Db	481	tcctttgactcttaacaagcagcgcctgtgtgtacacagaagaagtgagccactcttctgtgag	540
QY	541	gtccctcagctatgctcaagttactctgtttgttgagaacaacagaaatgagctgtgaactatc	600
Db	541	gtgcctcagactcttccaagtgtgtcatgtgtgacacaagaagctatgttggtgcgtgagctctc	600
QY	601	cttgtagtgagcctctccatctataatcccttgagcactatccctataatatatgtcttt	660
Db	601	ttcttttagtgtaacaattctctctataatccacagtgacatgtgaccccatctccctgaagctc	660
QY	661	atgtctcgagagcatgtgagatatacagttctgtgaagctgtgacaaaagaacatlttgtagaca	720
Db	661	ataagctcaagaatcttaaaaatcagttcagcagagaagcagcacaagaagcatlttgtagaca	720
QY	721	tgtgtgtcccatctaatgtgtgtgtctcttttatatgataagcgcgtctctgtgtactgt	780
Db	721	tgtgtgtcccaacagatgtgtgtgtgtccctcttttatatgaaacacgatlttatgtatctt	780
QY	781	caacacaccttgcgcagcctccaagaagacaaagaaagaatgtgtctctctctctatgtgactc	840
Db	781	caacacaccttcatcactccttaagaagatgtgtggaaagatgtgttccctctctatgtgactc	840
QY	841	atgcagcccatgtctgaatccctctatatatacaactatgtagaacaagaggttaaaagaaagc	900
Db	841	atgcagcccatgtctgaatccctctatatatacaactatgtagaacaagaggttaaaagaaagc	900

Dd 8A1 atcaatccacgtttgnaactcccctatctcacagccttcgaataaagatatgaaggagcc 900

Dy 901 tttaaaagtctggttgcagaagctctctctaacaagaa 939  
||| ||| ||| | ||||| ||||| ||| |||||  
Dd 901 ttcagaagcgctgatgccagaatctctttctgtaagaa 939

RESULT	9
AAH31645	
ID	AAH31645 standard; DNA; 936 BP

DT 30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 218.

KW Human; olfactory receptor; OR; primary scent determination;  
KW secondary scent determination; polypeptide library; odour receptor;  
KW scent profile; scent fingerprint; scent representation; ds.

OS Homo sapiens.

PN WO200127158-A2.

19-APR-2001.

PF 06-OCT-2000; 2000WO-US27582  
XY

PR	08-OCT-1999;	99US-0158615
PR	24-FEB-2000;	2000US-0184800

XX  
XX  
(Dict.) DISTINGUISH

PA (YEDA ) YEDA RES & DEV CO LTD.  
XX

PI Bellenson J, Smith D, Lancet  
xy

DR WPL; 2001-290713/30.

PT New polynucleotides  
PT sensation for identifi-

XX  
PS  
Claim 8. Page 263. 1857nn. English

XX  
XX  
The present sentence is one of a

The present sequence is one of a number of isolated polynucleotides which encode polypeptides involved in olfactory sensation. The polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. CC This enables the construction of a scent representation (also called CC a scent fingerprint or scent profile), which may be used to re-create CC and edit scents. Libraries of olfactory receptors are useful for CC determining the interaction pattern of a composition with the receptors CC and can be used for determining differences in the olfactory faculties CC of different individuals.

\$Q\$ Sequence 936 BP; 222 A; 226 C; 188 G; 300 T; 0 other.

Query Match	42.5%	Score 400.4	DB 22	Length 936
Best Local Similarly	64.9%	Pred. No. 5.2e-113		
Matches 593; Conservative	0	Mismatches 321	Indels 0	Gaps 0

[illegible]



QY	121	ggcaatcggaccatattcttagtgcagcgcctggagccccaactctataccoccatglat	180
Db	121	gggacaacaacacatcttgattatatacctacttggagccacactctccactctatglat	180
QY	181	ttttcttcaacaatctatacactcctctgagatcttgattacacacacatgtaacgtccaca	240
Db	181	ttttcttccccaacctaagcttttggatctggttaacacaaoccgcatgtgtccacag	240
QY	241	atgctagttaattatgcaagcatcaagaaagtaatacagttatcgttgctgtgtagccag	300
Db	241	ctctcgtgttaactcacaaggagacagacaacatccatccctatgtygttgtagtgcag	300
QY	301	ctttccatattttctggccttgaggagacacagtaaatcttccctggccgcacgtccctt	360
Db	301	ctgtacatctctcccaagctctggagctcacaagagcgttcttcttaaggatgtagtgaatt	360
QY	361	gataggttgtagtatacttgtagccctccctccatctaccagttlcaagacacagagatc	420
Db	361	gaccgcatgagagtggttttggaagccccctccatcaacagtagtcaagacccctgtctg	420
QY	421	tgacctcagttggcagccgcacatccctggtgttaactggtttttagtaacctagtgtygt	480
Db	421	tatgtcgtgagtggtcttactctatggtgcatgtgttttgccaactccattgcagagc	480
QY	481	acctgcactccacagccgcgcacatctggaaccccatgtagtagatacacttctctgaa	540
Db	481	gtgcacatctgtctttaaacccttctgagaaataaataagacacttctcttgtag	540
QY	541	gtccctgcactgctcaagttatcttggttgagacacaacagaaatgaggtcgaacttc	600
Db	541	gtccctccatgtctcaagcttgctggttggaacctactatgtaatgatactgaaccttc	600
QY	601	ctgtgcagtgagctctccatctataacccctgacactatccctatatataatgcttt	660
Db	601	tttctcagttgcatattcttctctgtacacctgttgcatatacatctcctatagtcag	660
QY	661	attgtccgaagatgtgaagatagatcgtctgtaagctgcagacaacaaagcatitggaca	720
Db	661	attgtcagggagatcagagataaagttagcaacagagcagagaaaagtgttgagaca	720
QY	721	tgtgtgtcccatctaatgtgtgtgtctcttttattatgatacagccgctcgtgtactg	780
Db	721	tgtgtgtcccatccatcacagtggtgttccctgttactagcgaagatatactgcttactc	780
QY	781	caaccaccttgcgccagctccaagagaccaagaaagtgttctctctctatggaatc	840
Db	781	cagcccggaacaacactctctcagagatcagggcagagttcatctctctctacacatc	840
QY	841	attgcacccatgctggaatccctctatataatacttgtagaacaagaggtfaaaggaagc	900
Db	841	attacacccatgataccaaccccttatataatacacttgagacaagaagtgtgaaagagca	900
QY	901	tttaaaagtgtgt 914	
Db	901	cttaagaagtgct 914	
RESULT 10			
AAH32063			
ID AAH32063 standard; DNA; 1002 BP.			
AAH32063:			
AC	XX	AAH32063:	
XX	XX	30-JUL-2001 (first entry)	
DE	XX	Human olfactory receptor polynucleotide, SEQ ID NO: 636.	
XX	XX	Human olfactory receptor polynucleotide, SEQ ID NO: 636.	
KW	XX	Human olfactory receptor. OR: primary scent determination;	
KW	XX	secondary scent determination; polypeptide library; odour receptor;	
KW	XX	scent profile; scent fingerprint; scent representation; ds.	
OS	XX	Homo sapiens.	

```

EN XM WO200127158-A2.
XX PD
XX PD 19-APR-2001.
XX PF
XX PR 06-OCT-2000; 2000WO-US27582.
XX PR 08-OCT-1999; 99US-0158615.
XX PR 24-FEB-2000; 2000US-0184809.
XX PA (DIGI-) DIGISCENTS.
XX PA (YEDA ) YEDA RES & DEV CO LTD.
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX DR WPI; 2001-290713/30.
PT New polynucleotides which encode polypeptides involved in olfactory
PS sensation for identifying olfactory agonists and antagonists -
PS Claim 8; Page 435-436; 1857pp; English.
XX XX
XX CC The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.
XX XX
XX SO Sequence 1002 BP; 210 A; 266 C; 217 G; 309 T; 0 other;
XX
XX
XX Query Match 39.3%; Score 370; DB 22; Length 1002:
XX Best Local Similarity 62.4%; Pred. NO. 1.2e-103;
XX Matches 580; Conservative 0; Mismatches 350; Indels 0; Gaps 0;
XX
QY 13 aatgacagcatatcaagaagttaattctgcgggtttctcagcacgactcgctgtag 72
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
qy 73 aatgagagacaacctgacgggttcataccctttaaagggtttctgatatactcagttacg 132
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 73 ttccaactccttggtctcttgaattcttacactgtgacacatcttgggaatcagacc 132
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 133 aaggtcataattgctcataltgtatgtctcgtatttactaacattttgggaataccacc 192
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 133 attattctagtgtaacgcttggaaccaactatacccccatgtatttttcttacc 192
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 193 atcattctcggttctcgtcttggaaccaagaagcttatatgacgtatattcttcctt 252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 193 aatcatacaacttcggaatcttgyttaacacacatatgacagtcaccaaagttagttaat 252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 253 catctcctcttccttgtaacgctgttcaacacagtgatattcccacgctctggtaaac 312
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 253 ttatgcagatcaagaagaaglaacagtatctgtgctggttagaccagcttttcatatt 312
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 313 ctgttggaacccaatgaataactatgcgcatagtgtgctgtgttgtaaccttaacaacctc 372
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 313 ctgcgcttgagggtcagtaaatctctctcgcgcgtcatgcttcttgatagtgttgta 372
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 373 catgcacctggagatccactgtagtgcgtccgccgggtgtgattctctgttaaccgtatg 432
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 373 gctatttctggccctctcatcatcaagttatcagtgacacagagactcgtcctcaagtt 432
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 433 gctgctctgcgcgtccctctccatcaactgtgtcctaagacataccatcctcgcagcttg 492
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 433 gcagcgcgatactcttggttactgtttttagtaactcagttgtgttltctaccctgacttc 492

```

```

CC The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.
XX
XX Sequence 1011 BP; 217 A; 252 C; 202 G; 340 T; 0 other;
S0
Query Match      38.7%   Score 364.2; DB 22; Length 1011;
Best Local Similarity 62.9%; Pred. No. 7.3e-102;
Matches 564; Conservative 0; Mismatches 333; Indels 0; Gaps 0;
QY 12 aaatgacagcatcactacagggaattctatcttgcgtttccagatcgacctggcgcga 71
    ||||| | | | | | ||||| | | | | | | | | | | | | | | | | | | | |
Db 9 aaatgcagaagtttgaaagccttccttatctactctcgtgatcttaacgycctcatctgcga 68
    ||||| | | | | | ||||| | | | | | | | | | | | | | | | | | | | |
QY 72 gtttcacctctgtgtgtctctcttgatctcttaccagtgtgacacatcttggaaactgcac 131
    | ||| ||||| | | | | | ||| | | | | | | | | | | | | | | | |
Db 69 agtagtgtctcttctgtgtatctatctactctactgtataacagtagatagaacaactgtt 128
    ||||| | | | | | ||||| | | | | | | | | | | | | | | | | | | | |
QY 132 catatttcatgttgcagcctcgagacccaactcatacccccatgtattttttcttac 191
    ||| | | | ||||| | | | | | | | | | | | | | | | | | | | | | |
Db 129 cactcatccctgtcatcacccggaactcccactccacactccacatgtaactctctcttc 188

```

Query Match	38.7%	Score 364.2	DB 22	Length 1011
Best Local Similarity	62.9%	Pred. No. 7.3e-102		
Matches 564	Conservative 0	Mismatches 333	Indels 0	Gaps 0
QY 12	aaatgacagacatcacagagatctatctcgtgtgtcttcacagatcgacctgctgga	71		
Db 9	aaatgcaagttttgaagctctctatcttacttacttgatctttcttaactgacctcatcga	68		
QY 72	gtttccacctgtgtgtctctctctgattctcttaacactgtgcaaccltttggaaatcgac	131		
Db 69	agtagttctctcttctgttactctgattctctactcttgaataacactgataagaaacctgt	128		
QY 132	catattcttagtgcgcgtgggacacaaactcttaacccccctgatatctttctctac	191		
Db 129	catctcatccctgtactaacctcgagctcccactcccaactcccaatgtaactcttccttc	188		
QY 192	caatctacactctctgatactcttctttaaaccacatgtacagttcccaacaatgctagtaa	251		
Db 189	aaatctccatcttctgagctctctgctgaacacacacagactctatctccctcagttgcg	248		
QY 252	tttaabgcagcatcaggaagaatcaatcagttacgtgtgctgtgtgtagccagctttcataat	311		
Db 249	tctctgvgcccggaanaagaccatctcttaactgtgtgtgtacagttcaacttacttgt	308		
QY 312	ctctggccttgggggctactatgatatctctctctggtccgtacatgtccctttagattgtc	371		
Db 309	tctctgacacttgggaacccgagagatgtgtctctactgttggtagtgccttagatctgtatgc	368		
QY 372	agcatttgtctggcctcccatctactccagttatcatgatgacaggaactctgacctccagtt	431		
Db 369	agcgtgtgtgtagaactcttgacatctacactgctctcctatgacacctcgtttctgctgctgtt	428		
QY 432	ggcagccgacatccctgtgttactctgttttagtaaacatcagttgtgtctgttactaacggtact	491		
Db 429	ggctggcgctctcttgggtgaagtgttttaacaacctcagcaactcatctcccttactctt	488		
QY 492	ccagctctgacctctgttgaccctccatgtgataatgataactttctctgtgaaagctccgtgact	551		
Db 489	cttgataaccctatgtagacacatcgctcttagttagtaactcttctctgtgaaagctccagact	548		
QY 552	gctcaagtatctctgtgtgttagacacaacagcgaatgaggtcgaactatctccttgcagtgta	611		
Db 549	tctctgacatcacatgtgtttagtaaccttaggcgaatgagctgtaaccctcatgycatgtaagctc	608		
QY 612	gctcttccatctaaaccccgacacatccatctatcatatctcttattgttgcagagc	671		
Db 609	catctttgtcttcaataacctctcaatcccatcttccatctctccatgtgcatatgcccgggc	668		
QY 672	agatattgaggaatacagtcctgtcgaaggctcgacaaaagaacatcttggagacatgtgtctcca	731		
Db 669	tgtaactgagcatgtcaataacacactgtgcttcagaaagtgcttagagacatgtgtgagccca	728		
QY 732	tctaattgtgtgtctctctttttatagtaagacggctctctgtgttactgtcaaacacactct	791		





CC which encode polypeptides involved in olfactory sensation. The  
CC polynucleotides can be used in screening for olfactory agonists and  
CC antagonists. The methods allow for the determination of primary  
CC scents and the identification of the odour receptors used to detect  
CC these primary scents. The methods also enable determination of  
CC secondary scents and the identification of combinations of odour  
CC receptors that are involved in detecting such secondary scents.  
CC This enables the construction of a scent representation (also called  
CC a scent fingerprint or scent profile), which may be used to re-create  
CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors,  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.  
XX  
XX Sequence 927 BP; 209 A; 253 C; 176 G; 289 T; 0 other;

Query Match 37.6%; Score 354; DB 22; Length 927;  
Best Local Similarity 62.1%; Pred. No. 9,6e-99;  
Matches 558; Conservative 0; Mismatches 340; Indels 0; Gaps 0;

QY 1 atgaattggtaaatgacagcatcatcagaagattattctgctgggtttcgaatcga 60  
DB 1 atgggattggcaatgagatccctaatgattcattcattccttagcttcacagacac 60  
QY 61 ccttgagctgagttccactcctctgtgtcttctgattcttcttaccgtgacacttt 120  
DB 61 cctgctcgagagctgtctctcttctgtattctgcttcttcttaccctccctctgtg 120  
QY 121 ggaacatcgacatattctagtgacagcctgagacacaaactcatacccatgatt 180  
DB 121 ggaacatccacataatcatcatcatcatctgatactccctctcattacacccatgttac 180  
QY 181 tttttcttaccatctcatcactcctgtgattctgtttacacacatgtaacagtcacaa 240  
DB 181 tttttctcagaacctctcttactgacatctgtcttactactacagcttctctccag 240  
QY 241 atgcatgaatttatgacatcgacatggaagtaacatcagttatcgctgctgttaccag 300  
DB 241 acctagtttaacttgcaagacacaaagacagacatcactaagggtgtgtgtgagcaca 300  
QY 301 ctcttcataattctggccttgagggtactgaatactctctctctggccgtctgctctt 360  
DB 301 ctctataattctctggcactgggctcactgaattatctctcttggcgtacagtgctctg 360  
QY 361 gatagatttgatgattcttgctgagccttccatcactaactaactgaacacagagactc 420  
DB 361 gatcggtacatgtgctgtcgaacacctccactatgtatgcatcatgaacccacagctt 420  
QY 421 tggcctcaggttgagcagcagatcctgggttactgttttagtaactaagtgtgtgtct 480  
DB 421 tgcacacagcgtgacatcatctccggccaggtgtgtgagttcccttaacatgata 480  
QY 481 accctgactccacagctgcactcgttgacccctatgtgatagatcacttctcgtgaa 540  
DB 481 actttactcctgaattgtcctctgtgtgacacatagagcttgaccatttatttggcga 540  
QY 541 gtcctcagctgctcaagtattctgtgttgagacacagcaatgagagctgaactatc 600  
DB 541 gtaccagactcttcaagtgtgtgtgtgtgagacacacatgtaactaagtgtgtctt 600  
QY 601 ctgtcagttgagctcttccatcattatccctcgacatcattcattatcattctttt 660  
DB 601 gtgttaagtgtctgttctgttcttccatccacagacattcattcattcattgtcttc 660  
QY 661 attgtcagagcagattatgagatcagctgtgtaaggtcgacacaaagacatttggaga 720  
DB 661 ataactcaagctgtgtctgagatcaatcagtagaggaagcacaagacattcagcagc 720  
QY 721 tgtgttccatcattaatgtgtgtgtctctttttatagaaagcggctgtctgttaccg 780  
DB 721 tgcctccacactacagtggtgtgtattatctatgacacataatcattcagtgtaacctg 780

QY 781 caacacacttgccagcagccacaggaacaaagatgttctctctcatgaaatc 840  
DB 781 caactagtgaagctagtagccacaggaaggaattatctctcttcaaccatg 840  
QY 841 attgaccacttgcatacccttatatacacttaggaacaaagatgaagaa 898  
DB 841 gtgaccacacttaactatcatctactacttaactaagaacaaagatgaagaa 898

## RESULT 15

AAH32045  
ID AAH32045 standard; DNA; 936 BP.

AC AAH32045;

BT 30-JUL-2001 (first entry)

DE Human olfactory receptor polynucleotide, SEQ ID NO: 618.

KW Human: olfactory receptor; OR: primary scent determination;  
secondary scent determination; polypeptide library; odour receptor;

KW scent profile; scent fingerprint; scent representation; ds.

OS Homo sapiens.

XX MO200127158-A2.

XX 19-APR-2001.

XX 06-OCT-2000; 2000WO-0527582.

XX 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

XX (YEDA) YEDA RES & DEV CO LTD.

XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI: 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory

XX sensation for identifying olfactory agonists and antagonists -

XX Claim 8; Page 428-429; 1857pp; English.

CC The present sequence is one of a number of isolated polynucleotides  
CC which encode polypeptides involved in olfactory sensation. The  
CC polynucleotides can be used in screening for olfactory agonists and  
CC antagonists. The methods allow for the determination of primary  
CC scents and the identification of the odour receptors used to detect  
CC these primary scents. The methods also enable determination of  
CC secondary scents and the identification of combinations of odour  
CC receptors that are involved in detecting such secondary scents.  
CC This enables the construction of a scent representation (also called  
CC a scent fingerprint or scent profile), which may be used to re-create  
CC and edit scents. Libraries of olfactory receptors are useful for  
CC determining the interaction pattern of a composition with the receptors,  
CC and can be used for determining differences in the olfactory faculties  
CC of different individuals.  
XX  
XX Sequence 936 BP; 180 A; 273 C; 243 G; 240 T; 0 other;

Query Match 37.4%; Score 352.4; DB 22; Length 936;  
Best Local Similarity 61.4%; Pred. No. 3e-98;  
Matches 566; Conservative 0; Mismatches 356; Indels 0; Gaps 0;

QY 1 atgaattggtaaatgacagcatcatcagaagattattctgctgggtttcgaatcga 60  
DB 1 atgggattggcaatgagatccctaatgattcattcattccttagcttcacagacac 60  
QY 61 ccttgagctgagttccactcctctgtgtcttctgattcttcttaccgtgacacttt 120

```
Db 61 cccagctgagatgattctttatagccalcctctctctatttgctgacctact 120
QY 121 ggcattctgacattatcttagtgcagcctggaaaccaacttcatacccccagtat 180
Db 121 gggaaactcaaccatcatctcttcccgctggagcccggtctcatcacccatgtaac 180
QY 181 ttttcttaccatctcatctccctgagctcttcttcttaccacacatgtaagtcaccacaa 240
Db 181 tcttccctccagcaacctctccctcttgagccttgcttctgctactagttcaatcccccaa 240
QY 241 atgctagtaaaattatgcaagcatcaggaagtaacatcagttacgctgctgtagccag 300
Db 241 atgctagtaaaattatgcaagcatcaggaagtaacatcagttacgctgctgtagccag 300
QY 301 ctttcataattctgacctggggcctgactgaatatctctccctggccgtaatgctctt 360
Db 301 cctcatgctctcttgcttggtggggccacagatgcatctgctggtgtagtgatct 360
QY 361 gataggtttagctatttctgagcctctccatctactcaatgtaacatgcaagagatc 420
Db 361 gacccgtaagtggcaggtgctggccgccccctccgctacacccgcatcagaaacccagctc 420
QY 421 tgcctccagttgcaagccgcatccctgggttaactggttttagtaactcagtggtgtct 480
Db 421 tgcctccagttgcaagccgcatccctgggttaactggttttagtaactcagtggtgtct 480
QY 481 accctgactctccagctgacctctgtgagccctatgtatagatacatttctctgtgaa 540
Db 481 accatcactctgcaactcccatgtgtggtgacccgaggtgagggatcttccctctgag 540
QY 541 gtcctgcaactgcaagtatctgtgtgtgtgagacaacagcaaatgaggtgaactatc 600
Db 541 gtcctgcaactgcaagtatctgtgtgtgtgagacaacagcaaatgaggtgtgtcaat 600
QY 601 ctgtcagtgagctctctccatcctaaccctgagacacacacacacacacacacacacac 660
Db 601 ggtgtcagtgagctctctccatcctaaccctgagacacacacacacacacacacacacac 660
QY 661 atgtccgagcagatattgagagatacagctctgtgaaggtcgaacaaagaatttgggaca 720
Db 661 atgtccgagcagatattgagagatacagctctgtgaaggtcgaacaaagaatttgggaca 720
QY 721 tgtgttcccatcattgtgtgtgtgtgtgtgtgttttattagtagacagcgtctctgttacctg 780
Db 721 tgtgttcccatcattgtgtgtgtgtgtgtgtgttttattagtagacagcgtctctgttacctg 780
QY 781 caaccaccttgcacagctcgaagcgaagcaaggaagatgttctctctctctctctctctc 840
Db 781 ctccgccaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 840
QY 841 atgcaaccatgtcgaatctcccttatataatacattaggaacaagaaggaaggaaggaagc 900
Db 841 atgcaaccatgtcgaatctcccttatataatacattaggaacaagaaggaaggaaggaagc 900
QY 901 tttaaaggttgttgcagaag 922
Db 901 ctgaggaaggttctgtgggaaag 922
```

Search completed: February 26, 2002, 05:05:19  
Job time: 3117 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 26, 2002, 03:41:21 ; Search time 45.65 Seconds  
(without alignments) 4673.434 Million cell updates/sec

Title: US-09-755-017-1

Perfect score: 942

Sequence: 1 atgaattgggttaatgacag.....tcttcttaatcaagaataaa 942

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCRNUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216.8	23.0	1290	2	US-08-827-291A-1
2	211.6	22.5	966	3	US-08-748-506-7
3	211	22.4	966	3	US-08-748-506-8
4	200.8	21.3	1713	2	US-08-467-948A-1
5	200.8	21.3	1713	2	US-08-467-948A-1
6	200	21.2	900	4	US-09-085-371-5
7	197.2	20.9	966	3	US-08-748-506-5
8	190.4	20.2	966	3	US-08-748-506-6
9	183.6	19.5	984	3	US-08-748-506-9
10	118.4	12.6	1474	1	US-08-465-980-1
11	118.4	12.6	1474	2	US-09-053-303-1
12	118.4	12.6	1474	5	PCT-US95-07093-1
13	94.4	10.0	1828	3	US-08-988-876-2
14	90.2	9.6	1320	1	US-08-599-252-84
15	90.2	9.6	1320	1	US-08-436-074-57
16	90.2	9.6	1320	5	PCT-US96-06352-84
17	90.2	9.6	1320	5	PCT-US96-06352-84
18	57.8	6.1	7210	5	PCT-US96-06583-84
19	41.6	4.4	1260	1	US-08-232-463-14
20	41.4	4.4	1260	1	US-08-599-252-83
21	41.4	4.4	1260	1	US-08-436-074-56
22	41.4	4.4	1260	5	PCT-US96-06352-83
23	41.4	4.4	1260	5	PCT-US96-06583-83
24	39.8	4.2	2012	3	US-08-706-281A-7
25	39.8	4.2	2012	4	US-09-097-231-7
26	36.6	3.9	1015	1	US-08-671-525B-3
27	36.6	3.9	1015	1	US-08-672-109B-3

28	36.6	3.9	1015	1	US-08-842-045-3	Sequence 3, Appl1
29	36.6	3.9	1015	2	US-08-842-238-3	Sequence 3, Appl1
30	36.6	3.9	1015	3	US-08-629-338B-3	Sequence 3, Appl1
31	36.6	3.9	2012	1	US-07-866-560-3	Sequence 3, Appl1
32	36.6	3.9	2012	1	US-08-077-673-3	Sequence 3, Appl1
33	36.6	3.9	2012	1	US-08-478-992-3	Sequence 3, Appl1
34	36.6	3.9	2012	1	US-09-105-298-3	Sequence 3, Appl1
35	34	3.6	289	4	US-09-007-005-17	Sequence 17, Appl1
36	34	3.6	289	4	US-09-244-796-17	Sequence 17, Appl1
37	32.4	3.4	753	4	US-08-998-416-574	Sequence 574, App
38	31.8	3.4	1101	3	US-08-945-056-5	Sequence 5, Appl1
39	30.4	3.2	603	3	US-08-609-334-10	Sequence 10, Appl1
40	30.4	3.2	603	3	US-08-609-334-15	Sequence 15, Appl1
41	30.4	3.2	1092	4	US-09-077-675A-15	Sequence 15, Appl1
42	30.4	3.2	3129	4	US-09-077-675A-14	Sequence 14, Appl1
43	30.2	3.2	1101	3	US-08-945-056-7	Sequence 7, Appl1
44	30.2	3.2	80246	4	US-09-078-294-4	Sequence 4, Appl1
45	30.2	3.2	80595	4	US-09-078-294-3	Sequence 3, Appl1

#### ALIGNMENTS

RESULT 1

US-08-827-291A-1

Sequence 1, Application US/08827291A

Patent No. 5874243

GENERAL INFORMATION:

APPLICANT: Macina, Roberto

APPLICANT: Sathe, Ganesh

TITLE OF INVENTION: NOVEL OLRCCL5 RECEPTOR

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Smithline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY:

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/827,291A

FILING DATE: 28-MAR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: King, William T

REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: GP50001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5015

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1290 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-827-291A-1

Query Match 23.0%; Score 216.8; DB 2; Length 1290;

Best Local Similarity 52.2%; Pred. No. 4.7e-60;

Matches 482; Conservative 0; Mismatches 442; Indels 0; Gaps 0;

```

OY 1 atgaattggtaataagacagacatcacaggaattatctgtggttctccaaatcga 60
DB 296 ATGGCAGGAGAAATTCGACCTTCACACCTTCATCTTCTCGGATCTTTCATCAC 355
OY 61 ccttgctgagatccacacccctctgtgtctcttgattcttaacatgacacatctt 120
DB 356 AGCCCCACCCACACCTTCCTCTTCTTCTGTGCTGCGCATCTTTCATGCGCTTCG 415
OY 121 ggcacatctgacccatctatctatgtacagcctggagacacacacacacacacac 180
DB 416 GGAACCTCTGTCATGGTCTTCTCTCATCTACCTGGACACCCACCTCCACACCCCA 475
OY 181 tcttcttaccacatctacactccctcgtgacacacacacacacacacacacacacac 240
DB 476 CTCCTCTCAGCCAACTGTCCTCATATGACACCTCATGCTGACACACCCGATACCA 535
OY 241 atgctagtaaatctatgacagacatcaggaagaatcaatcagtaatcgtgtgtgac 300
DB 536 ATGGCCTTCACACTCTGCTGCGACAGATCCATTTCTATGGCTGTGCTGCGACACA 595
OY 301 ctttcatatctctgctggggtggtgactgaaatactctctcctgacgctgacac 360
DB 596 ATTTCTTTTATACATCACTGCTGCTGCTGATGCTTCTTCTTGTGCTTATGCTTAT 655
OY 361 gataaggttctagctatctgtcgcctccatctacacagtaatactgacacagagac 420
DB 656 GACCGGTACACCTGATGCTGCTGACACCTCATACATACACCACTCATGACCCCTAA 715
OY 421 tgcctccagctggcagcgcacccctggtgtactggttttagtaactcagtggtgt 480
DB 716 TGTGACTTATGACTGCTCTTCTGCTGATCTGCGCTCTACAGATGAGATCAATTTAT 775
OY 481 acccgcactcagctgcacactcctgtgacccctatgtatgatactatctctctgtga 540
DB 776 GTAGCCACATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 835
OY 541 gtcctgcacgtcacaagtatctgtgtgtgacacacacacacacacacacacacacac 600
DB 836 TTACCTTCCCTACTATCTCTCATGACATGACACATCAATTTGAAAGGTTATTTTC 895
OY 601 ctgtgcagtgagctctctcatctaatcaccctgacacatctctctatacatgtctt 660
DB 896 ATTGCTCTATATGATATGCTTGTCTTCTCTGTCATCATATGCTTCTCTATGCTGA 955
OY 661 attgctgcagtgagctctgtgacagctgacagcagacacacacacacacacacacac 720
DB 956 GTTATTTCTGCTGCTCATTCATACATGGGATGTGAGAGGCTGCTGCAAACTTTCA 1015
OY 721 tglgttcccatctaatgtgtgtgtctctttttagtagacagcgctcgtgtgtac 780
DB 1016 TGTCTCTCTCACCCTCATGCTGCTGGAATGCTTCTATGAGCAGGTTGTTCA 1075
OY 781 caacacacttgcacacacacacacacacacacacacacacacacacacacacacac 840
DB 1076 CAGCCACACTGTGATGCTGCTCCCAACGACGACAGCTGCTGCTGATTTTCA 1135
OY 841 attgacacacacacacacacacacacacacacacacacacacacacacacacacac 900
DB 1136 CTCACCTCCATGCTCAATCCCTCATCTACACCTCCGACACAGAACTGACACAGCA 1195
OY 901 tttaaaaggttggttgcaagagtc 924
DB 1196 TTCATGAGATCTCAGAAAGGCG 1219

```

# RESULT 2

US-08-748-506-7  
 : Sequence 7, Application US/08748506  
 : Patent No. 6159707  
 : GENERAL INFORMATION:  
 : APPLICANT: Bonnett et al.  
 : TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
 : NUMBER OF SEQUENCES: 31

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Volt & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/748,506
FILING DATE: 08-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,751
FILING DATE: 09-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 74940
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5600
TELEFAX: 312-616-5700
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-748-506-7

```

```

Query Match      22.5%  Score 211.6; DB 3; Length 966;
Best Local Similarity 55.2%  Pred. 1.9e-58;
Matches 412; Conservative 0; Mismatches 334; Indels 0; Gaps 0;

OY 163 cttaatacccatgatttttcttaccatcatcactcctgtgatttgttaacc 222
DB 181 CTACACACCCCATGACTTCTTCTGSCCACTGTCTCTCGAGATGGCTATTA 240
OY 223 acatgtacagctcccaaatctagtaaatctatgacatcagaagaatlaacgtat 282
DB 241 TGTCTGTGATACCAAGATGCTGCGACAGCTTGTGAGGAGGCCGAGGATCTT 300
OY 283 cgtgcgtgtgacacagctcttcatatcttgcacctggtgggtactgaatatcttc 342
DB 301 GAGGCTTGGCCCTACAGATGTTCTTCTCATATCTTGTGATTAACAGAGCTG 360
OY 343 ctggcgcatactctttagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 402
DB 361 TTGGCAGCCATGCTTGTACCGCTATATGCTATATGTTCCCACTCCATGCAACC 420
OY 403 atcatgacacagacacacacacacacacacacacacacacacacacacacacacac 462
DB 421 CCAATGAGCTGCGGATGATGCTGATTTGGCAATTTGCTCATGCTGATGAGATG 480
OY 463 aactcagtggtgtctacacacacacacacacacacacacacacacacacacacac 522
DB 481 GTAGCTGTGGACAGACCAATTTTATTTTCTCTGAACTTCTGAGACCTGTGATA 540
OY 523 gatcacttctctgtgaagctccctgacacacacacacacacacacacacacacacac 582
DB 541 GACACTTCTTGTGACCTTCTCACCTTCTGCGACCTTCTGCTGCTGATATCCCA 600
OY 583 aatgagctgaactatctctgtcagtgagccttccatcactaaccctgacacacac 642
DB 601 AATGAGCTGCAATCTTTTGTGCGACAGAGCTGCTGATTTTATGTCATTTTACTGA 660
OY 643 ctatacatatgtcttattgttcgagcagatgtgaagatacagctcgtgtgaagtcga 702

```



OY 163 ctcaatccccatgatatctttcttcaacaacatcatccatccgagatcttcttacc 222  
 Db 181 ctacacacccccatgacttcttctctgcccacacttgcctctctctcgagatggctact 240  
 OY 223 acatgtaacgtccccaacaatgctagtaattatgacgacatcsgaagaatcaatcagttat 282  
 Db 241 tgcctcttcataaccacgaatgcctccaaagctctgtgatgtagcccgagagatctttcag 300

```

1  COUNTRY: USA
2  ZIP: 20005
3
4  COMPUTER READABLE FORM:
5
6  MEDIUM TYPE: FLOPPY DISK
7
8  COMPUTER: IBM PC COMPATIBLE
9
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11
12 SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30

```



```

TITLE OF INVENTION: Vivo
FILE REFERENCE: P01511US2 / 09805059
CURRENT APPLICATION NUMBER: US/09/085,371
CURRENT FILING DATE: 1998-05-19
PRIOR APPLICATION NUMBER: US 08/891,243
PRIOR FILING DATE: 1997-07-10
PRIOR APPLICATION NUMBER: US 60/045,961
PRIOR FILING DATE: 1997-05-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 900
TYPE: DNA
ORGANISM: Rat
US-09-085-371-5

21.2%; Score 200; DB 4; Length 900;
Best Local Similarity 53.0%; Pred. No. 1e-54;
Matches 459; Conservative 0; Mismatches 399; Indels 12; Gaps 1;

```

[illegible]



Tue Feb 26 09:22:07 2002

us-09-755-017-1.rni

Page 7

```

? REFERENCE/DOCKET NUMBER: 7494-10
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 312-616-5600
? TELEFAX: 312-616-5700
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 966 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? OS-08-748-506-6

```

Query Match	20.28;	Score 190.4;	DB 3;	Length 966;
Best Local Similarity	53.4%;	Pred. No. 1.3e-51;		
Matches 398;	Conservative 0;	Mismatches 348;	Indels 0;	Gaps 0;

[illegible]

RESULT 9  
US-08-748-506-9

Sequence 9, Application US/08748506  
Patent No. 6159707

GENERAL INFORMATION:  
APPLICANT: Romnett et al.  
TITLE OF INVENTION: NOVEL Sperm RECEPTORS  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leydig, Volt & Mayer, Ltd.  
STREET: Two Prudential Plaza, Suite 4900  
CITY: Chicago

COUNTRY: US

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 ADDITION REQUESTED: YES (00 2740 5000)

FILING DATE: 08-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 80/033,731  
FILING DATE: 09-NOV-1985

CLASSIFICATION: 435

REFERENCE/DOCKET NUMBER: 74940

TELEPHONE: 312-616-5600

TELEFAX: 312-818-3700  
 INFORMATION FOR SEO ID NO: 9:

SEQUENCE CHARACTERISTICS:

TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

every match	19.58; Score 1
-------------	----------------

Category	Count	Percentage
Conservative	444	0%
Misplaced	0	0%

13 aatnaccatcataccanantfatctac

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]

91 ATCCCTCTTCTTCCCTACTGCACTTTGCTGGCCCT

133 attatctagtgtcaagcctggacaaccaaac

151 ATAATTACCATCACCTGTGTGGACCACCGAC

193 aatctatcactccctgqatcttqttaacacca

211 A C C T T C T T C T T T T G C G C A G T C G T T G T T T A T A A

הנה

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

313 ctggccltgggctaactggaataccttccccc

331 CTTGTCGTGGGGCAACAGTTTTCCTTT

373 gctattgtcgacctctcattactcagtta

391 GCCATCTGCAACCCTCTACATTATCCAACCP

Oy	433	gaagcgcgaatccctgggttaacggttttaagtaaaccaagtggtgtgtctacccggaactc	492
Db	451	gTTACTGCTGTTATTTATTTTGGGCTTCCTCTTCAATGGCCAGATCCAGTTGATGCTTTCC	510
Oy	493	cagctgcacactcgtlgaaacccatgtagatgaatcaacttctctgtgaagtcctcgaactg	552
Db	511	AAGACATTTTCTGTGGTCCAAACATTAATTCCTCACTTTTCTGTGATTTTGGAGCCATG	570
Oy	553	ctcaagatcatcttgyttagagacaacagcaatbagagctgaacatctcctgtgcagag	612
Db	571	GCAAATCTCTCTGTTTCAGAANAACAGGCTATTATAGATGCTGTTTTTACCTTTCGTGA	630
Oy	613	ctctcccaatcctaataccccctgacacatccatccatatacatgatttaatgtgcagaca	672
Db	631	ATTGGCTTTTGGCTTCCTTTCTTATAGCCATCTTTCGATACAGCAATATAGTAGTCAC	690
Oy	673	gtattgagatacagctcgtctgaaggtgcgacaacaagaacttggacatggtgttcccat	732
Db	691	ATAGGAGACATCCCTTCAGCCAGGAGGACAGAGAGACTTTTTCACCTCTCTCCAT	750
Oy	733	ctaattgcggtctcctctttttaaagaaagcgcgtctcgttgaacctgcgaacacactcg	792
Db	751	CTCATTTGCCCTCTCTCTATATATATAGGCAAGCTGTGCATTTATATCCCTGAAGCCAAAGAC	810
Oy	793	cccagctccaaagaccaaagaagaatggttctctcttctaagaaatcattgcaaccatg	852
Db	811	AGAACACAGAGTGACACCAACAGAGAGGCTCTGTTGAACAATGAGTGTGAGACACCCCTT	870
Oy	853	ctgaatcccttatataaacacttaagaaagaagagt	890
Db	871	CTGAACCTGTCAATCTACACCTCGGCAACAAGCAGGT	908

```

1      RESULT 10
2      US-08-465-980-1
3      ; Sequence 1, Application US/08465980
4      ; Patent No. 5756309
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Soppet, Daniel R.
7      ; APPLICANT: Li, Yi
8      ; APPLICANT: Rosen, Craig A.
9      ; APPLICANT: Ruben, Steven M.
10     ; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPR470
11     ; NUMBER OF SEQUENCES: 8
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: CAPELLA, BYRNE, BAIN, GILFILLAN, CECCHI
14     ; ADDRESSEE: START & OLSTEIN
15     ; STREET: 6 Becker Farm Road
16     ; CITY: Roseland
17     ; STATE: New Jersey
18     ; COUNTRY: USA
19     ; ZIP: 07068
20     ; COMPUTER READABLE FORM:
21     ; MEDIUM TYPE: Floppy disk
22     ; COMPUTER: IBM PC compatible
23     ; OPERATING SYSTEM: PC-DOS/MS-DOS
24     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
25     ; CURRENT APPLICATION DATA:
26     ; APPLICATION NUMBER: US/08/465,980
27     ; FILING DATE: 06-JUN-1995
28     ; CLASSIFICATION: 536
29     ; ATTORNEY/AGENT INFORMATION:
30     ; NAME: Ferraro, Gregory D.
31     ; REGISTRATION NUMBER: 36,134
32     ; REFERENCE/DOCKET NUMBER: 325800-446
33     ; TELECOMMUNICATION INFORMATION:
34     ; TELEPHONE: 201-994-1700
35     ; TELEFAX: 201-994-1744
36     ; INFORMATION FOR SEQ ID NO: 1:
37     ; SEQUENCE CHARACTERISTICS:
38     ; LENGTH: 1474 base pairs
39     ; TYPE: nucleic acid
40     ; STRANDEDNESS: single

```

```

?      TOPOLOGY:  linear
?      MOLECULE TYPE:  CDNA
?      FEATURE:
?      NAME/KEY:  CDS
?      LOCATION:  274..1233
US-08-465-980-1

```

Query Match	12.68;	Score 118.4;	DB 1;	Length 1474;
Best Local Similarity	48.98;	Pred. NO. 2.6e-28;		
Matches 317;	Conservative	0;	Mismatches 331;	Indels 0;
			Gaps	0;

QY	109	gtgacacattcttggaacatcgcaccattatcttagtgcacgcccgcagaccacaaattcat	168
Db	379	GTGGCAATGTTGGAAACGTGATGTGTGTTCTTATCGTAAGACGAGAACGACCTTCAC	438
QY	169	accgccatgtaatttttcttaccaatctacacctgagctcttgattacaccaaagt	228
Db	439	GCTCGGATGTCACCTTTTCTTCATCGATCTTGACACCATTGACCGTGCGCTTATCCATTC	498
QY	229	acagtcaccacaatgctagtaaatattatgcagcctcaggaagtatacgaattatcgtgyc	288
Db	499	ACCAATGCTTAAGATCTTGCCCTTTCTGTGGTTGATTCCGAGAGATTACATTGAAGCC	558
QY	289	tgtgtgacccagctttcataattctctgaccttgaggcactagatactctccctggcc	348
Db	559	TGTCTTAOCCAGATGTTCTTTATTCAAGCCCTCTACACCAATTGAATGACACATCTCTG	618
QY	349	gtcaatgctccttgtaagtttgtagctatttgcgcctctccatcaattcagttatcatg	408
Db	619	GCCATTGGCCTTTGACCCGTTAATGTGGCCATCTGCCACCCACTGGCCCATGCTGGAGTGC	678
QY	409	caccagagaccctgcctccagttgagcagccgacatcccggtttacgtgttttagtaactca	468
Db	679	AACATTAAGTAACAGCCCAATGGATGGATCGTGCTGTGGTCGGGATCCCTCTTTT	738
QY	469	gtgtgtgtgtctacccgaactcctccagctgcacactctgtgacccctatgtaataatc	528
Db	739	TTTCCACAGCCTCTGCTGATCAACGGGCTGGCCTTTCGCCATTCGATGTCCTGTGCAC	798
QY	529	tttctctgtgtaagtcctgtcaactgactaaagtatctctgtgtgtgaagacacagcaaatgag	588
Db	799	TCTCATTTGTGCCACCGAGATGATGAAGTTGGCCATTCAGACACTTTGGCCAAATGTG	858
QY	589	gctgaactatctctgtgtaagtgagctcttcacatctaatatcccccagacacatccattata	648
Db	859	GTAATAGTCTTACTGCCATTCTCTGTGTCAATGGGCGTGGACGTAATGTTCAATCTCTTG	918
QY	649	tcaatagcttattatgttcgcagacgatatgagaatacagctctgcgtgaagtgacaanaa	708
Db	919	TGCTAATTTTCTGATAATACGAACGGTCTCTGCACATGCCCTTCCAAGTACAGAGCGGGCCAG	978
QY	709	gcatttggagatatglttccacatcctaattgtgtgtctcttttatt	756
Db	979	GCCTTTGAACCTGTGTCTACCAATTGTGTGTACTCGCCTTCTAT	1026

RESULT 11  
 US-09-053-303-1  
 : Sequence 1, Application US/09053303  
 : Patent No. 5948890  
 : GENERAL INFORMATION:  
 : APPLICANT: Soppet, Daniel R.  
 : APPLICANT: LI, YI  
 : APPLICANT: Rosen, Craig A.  
 : APPLICANT: Ruben, Steven M.  
 : TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPRA170  
 : NUMBER OF SEQUENCES: 8  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: CAELLA, BYRNE, BAIN, GILFILLAN, CECCHI  
 : ADDRESSEE: STUART & OLSTEIN  
 : STREET: 6 Becker Farm Road  
 : CITY: Roseland



```

OY 229 acagtcaccacaaatgctagtaaatatgacagcatcaggaagaatcagttacgtgac 288
DB 499 ACCATGCGCTTAAGATCCTTCCCTTTCTGCTTGAATCCCGAGAGATTAGCATTTGAGGCC 558
OY 289 tctgtgagccagcttctcatattctgacctggtgggagctacgtatattctcctgacc 348
DB 559 TCTCTTACCCAGATGTTCTTATTCATGCGCCCTCAGCAGCATGAAATCCACCATCCTGCGTG 618
OY 349 gtcacgtcttgataggtttgtagctatttggcgccctccatctacacgtatcagtg 408
DB 619 GCCATGCGCTTACCGCTTATATGCGCATCTGCCACCCCTGCCCATGCTCAGTGTCTC 678
OY 409 caccagagactctgctcccaagttggcagcgcatcctcgtgttacttggttttagtaactca 468
DB 679 AACATATACAGTAACAGCCAGATGCGCATGCGGCTGTGCGCGGATGATCCCTCTTTT 738
OY 469 ggtgtgtgtctcaacctgactctcagctgacactctgtgacctgtatgtatgatactac 528
DB 739 TTCCCACTGCTCTGCTGATCAAGCGGCGGCTTCTGCACTCCCATATGCTCTGCGAC 798
OY 529 tttctctgtgaagtcctcctcactgctcaagttatcttgtgtgagacaacagcaaatgag 588
DB 799 TCCATATGCTGTCACCAAGATGATGAAATGGCTTATGACAGACTTTGCCCATGTG 858
OY 589 gctgaactatctcctgtcagtgagctctcactcaataacccctgacactacatccttata 648
DB 859 GATATATGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 918
OY 649 tcatatgcttattgtgtcagcagtatgagatacagctgctcgtgaagtgatgacaaaa 708
DB 919 TCCATATGCTTATATATGAAAGGCTTCTGCACTGCTTCCAAAGTCAGAGCGGCGGCAAG 978
OY 709 gcaattggagacatgctgttcccatctaatattgtgtgtctcttttat 756
DB 979 GCCTTTGGAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1026

```

# RESULT 13

US-08-988-876-2

Sequence 2, Application US/08988876

Patent No. 6063596

GENERAL INFORMATION:

APPLICANT: Lal, Preeti

APPLICANT: Bandman, Olga

APPLICANT: Hillman, Jennifer L.

APPLICANT: Yue, Henry

TITLE OF INVENTION: G PROTEIN COUPLED RECEPTORS ASSOCIATED

TITLE OF INVENTION: WITH IMMUNE RESPONSE

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: PASCSEO for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/988, 876

FILING DATE: Herewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0441 US

TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1828 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNOT01
CLONE: 364702
US-08-988-876-2

```

```

Query Match 10.0% Score 94.4; DB 3; Length 1828;
Best Local Similarity 55.4%; Fred. No. 1.5e-20;
Matches 205; Conservative 0; Mismatches 161; Indels 4; Gaps 1;

```

```

OY 34 tttatctgctggttttccagatcgacctgctggaatttccactcctgtgtctc 93
DB 814 TTTCTTCTCTCTAGAGTCTCAGGGGATCCAGACTGCAAGCCAGTCTTGCTGGCTGTT 873
OY 94 ttgatttctaacctgtgacacatcttggcaactgacatattcttagtgacgctg 153
DB 874 CTGTCCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 933
OY 154 gacacaaactcataccccccatgtatttttcttaccatctacactcctgtatct 213
DB 934 GACTCCACCTCCACACCCCAATGATCTTCTCTCCAACTGCTCTTGCTGACATC 993
OY 214 tgtacacacatgtagagttccacacaatgctagtaattttagcagacacagaaga 273
DB 994 GGTTCACCTCCACACAGCGTCCCAAGATGATGATGATGATGATGATGATGATGATG 1053
OY 274 atcagttatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 332
DB 1054 ATCTCTATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1113
OY 333 ---atactctcctggtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 389
DB 1114 GAGAGACATGCTCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 1173
OY 390 ccatactca 399
DB 1174 ATATCATCA 1183

```

# RESULT 14

US-08-599-252-84/C

Sequence 84, Application US/08599252

Patent No. 5705343

GENERAL INFORMATION:

APPLICANT: DRAYNA, DENNIS T.

APPLICANT: FEDER, JOHN N.

APPLICANT: GRIEKE, ANDREAS

APPLICANT: KIMMEL, BRUCE E.

APPLICANT: THOMAS, WINSTON J.

APPLICANT: WOLFE, ROGER K.

TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY

TITLE OF INVENTION: HEMOCHROMATOSIS

NUMBER OF SEQUENCES: 124

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & ROENSTER

STREET: 2000 Pennsylvania Ave. N.W., Suite 5500

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS



```

: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/599,252
: FILING DATE: 09-FEB-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MURASHIGE, KATE H.
: REGISTRATION NUMBER: 29,959
: REFERENCE/DOCKET NUMBER: 9053-0001.21
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0763
: TELEX: 90-4030
: INFORMATION FOR SEQ ID NO: 84:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1320 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-599-252-84

```

Query Match	9.6%	Score 90.2	DB 1	Length 1320
Best Local Similarity	65.8%	Pred. No. 2.9e-19		
Matches 131	Conservative	0	Mismatches 66	Indels 0
QY 718	acatgctggtcccatcctaactcgtggtgcctcttllttagtagacagccgctcgtgtac			777
Db 1319	ACTTGCACCTGTGCATCTATTGTTCTCTCTCTTCTTGTGTACATCATATACATGTAC			1260
QY 778	ctgcacaccacctctgcgcagctcccaagagcacaagaagaatcgtttcctctcatgga			837
Db 1259	CTCCAGCCACGAGAAATACTTATTCCAGAGACCGGGGCAAGTTCTTACCTTTTTCACACA			1200
QY 838	alcattgcaccocatcgtcaatcccttatataataacttaggaacaagaaglaaagaa			897
Db 1199	ATTGTCACTCCACAGGTTAACCCCTGTATCTATACACTAAGAACAAACAAATTTAAAGAG			1140
QY 898	ggccttaaaaggttggttg	916		
Db 1139	GCCATGAAGAAAGTGCTGAG	1121		

RESULT 15  
US-08-436-074-57/c  
Sequence 57, Application US/08436074  
Patent No. 5753438  
GENERAL INFORMATION:  
APPLICANT: DRAVNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GNIKE, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFE, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
TITLE OF INVENTION: HEMOCHROMATOSIS  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,074  
FILING DATE: 08-MAY-1995  
CLASSIFICATION: 436

```

1 ATTORNEY/AGENT INFORMATION:
2 NAME: MURASHIGE KATE H
3 REGISTRATION NUMBER: 29,959
4 REFERENCE/DOCKET NUMBER: 9053-0001.00
5 TELECOMMUNICATION INFORMATION:
6 TELEPHONE: (202) 887-1500
7 TELEFAX: (202) 887-0763
8 TELEX: 90-4030
9 INFORMATION FOR SEQ ID NO: 57:
10 SEQUENCE CHARACTERISTICS:
11 LENGTH: 1320 base pairs
12 TYPE: nucleic acid
13 STRANDEDNESS: single
14 TOPOLOGY: linear
15 US-08-436-074-57

```

	Query Match	9.6%	Score 90.2;	DB 1;	Length 1320;	
	Best Local Similarity	65.8%;	Pred. No. 2.9e-19;			
	Matches 131; Conservative	0;	Mismatches 68;	Indels 0;	Gaps 0;	
OY	718 acatgttgattcccatctcaatgtytgytgcctttlltatagtaacggcgtctcytgytac	777				
Dd	1319 ACTTGCACCTGCCAATCTATTGTTGTCTCTCTCTATAGTACAATCATATATACGTATAC	1260				
OY	778 ctgcacaaccacctcgcccgcgaactcaagaagcaagaagaatgtytcttcctctatgta	837				
Dd	1259 CTCACAGCCACGAATAACTATTATTCACAGACACGGCGCAAGTTCTTACCCTTTTTCACACA	1200				
OY	838 atcaattgacccaatcvtgaatcccttatatatatacatcttaggaacaagaagttaaaggaa	897				
Dd	1199 ATTGTACACTCCCAGAGTTAAACCCCTTGATCTATACACTAGAACAACAAGATGTTTAAGAG	1140				
OY	898 ggcctttaaaaggttgtytgyt 916					
Dd	1139 GCCATGAAAGAGGTCTCTAG 1121					

```
Search completed: February 26, 2002, 05:02:56
Job time: 4895 sec
```



OM of: US-09-755-017-2 to: GenBank.\* out\_format: pfs

Date: Feb 26, 2002 5:52 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

# Command line parameters:

```
-MODEL=framed-p2n -DEV=xlh
-Q=/cgrn2.1/USFTO.spool/US09755017/runtat.25022002.165640.20772/app.query.fasta_1.374
-DB=GenBank -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OCGAPOP=4.500 -OCGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsun62
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR.SCORE=pct
-THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09755017.@CNCI_1.5603 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREDS=1
```

## Search information block:

```
Query: US-09-755-017-2
Query length: 313
Database: GenBank.*
Database sequences: 1472140
Database length: -341344837
Search time (sec): 1363.720000
```

## Score list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
gb_pr:AL133267	-	1601.00	2242.47	1.2e-116	44788	AL133267 Human DNA sequence
gb_hcg:AC025941	-	1601.00	2230.22	6.0e-116	176277	AC025941 Homo sapiens chrom
gb_pr:HS193612	-	1323.00	1879.97	1.9e-96	1074	AJ302593 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302584 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302585 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302586 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302587 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302588 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302589 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302590 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302591 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302592 Homo sapiens 6M1-10*0
gb_pr:HS193612	-	1319.00	1874.27	4.0e-96	1074	AJ302593 Homo sapiens 6M1-10*0
gb_hcg:AC025941	-	1319.00	1829.17	1.3e-94	100375	AC025941 Human DNA sequence
gb_hcg:AC025941	-	1319.00	1829.17	1.3e-94	166758	AC025941 Homo sapiens chrom
gb_hcg:AC025941	-	1319.00	1829.17	1.3e-94	176277	AC025941 Homo sapiens chrom
gb_hcg:AL589651	-	1317.00	1825.37	2.1e-93	185523	AL589651 Mus musculus chrom
gb_hcg:AL589651	-	1317.00	1825.37	2.1e-93	222824	AL589651 Mus musculus chrom
gb_hcg:AL589651	-	1302.00	1802.38	4.1e-92	222824	AL589651 Mus musculus chrom
gb_hcg:AL589651	-	1267.00	1799.13	6.2e-92	1214	L34074 Rat Olfactory receptor gene
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302537 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302538 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302539 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302540 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302541 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302542 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302543 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302544 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302545 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302546 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1138.00	1617.71	7.8e-82	942	AJ302547 Homo sapiens 6M1-10*0
gb_hcg:AL589651	-	1106.00	1579.48	1.8e-79	649	U86270 Homo sapiens Olfactory
gb_hcg:AL589651	-	1062.00	1512.84	5.4e-76	648	U86271 Homo sapiens Olfactory
gb_hcg:AL589651	-	1047.50	1441.64	5.0e-72	185302	AL359332 Mus musculus chrom
gb_hcg:AL589651	-	1047.50	1439.90	7.2e-72	251546	AL359336 Mus musculus chrom
gb_hcg:AL589651	-	1032.50	1470.86	1.2e-73	646	U86275 Homo sapiens Olfactory
gb_hcg:AL589651	-	955.00	1308.70	1.3e-64	212569	AL592522 Mus musculus chrom
gb_hcg:AL589651	-	955.00	1308.24	1.4e-64	223791	AC084042 Mus musculus chrom
gb_hcg:AL589651	-	955.00	1307.55	1.5e-64	241714	AC083834 Mus musculus chrom
gb_hcg:AL589651	-	949.00	1304.53	2.2e-64	130279	AL121944 Human DNA sequence
gb_hcg:AL589651	-	945.00	1297.44	5.4e-64	152311	AC011571 Homo sapiens clone

```
seq_name: gb_pr:AL133267
seq_documentation_block:
LOCUS AL133267 44788 bp DNA PRI 25-OCT-2000
DEFINITION Human DNA sequence from clone RP3-408B20 on chromosome 6 Contains
ESTs, STSs and GSSs. Contains a gene and two pseudogenes for novel
7 transmembrane receptors (olfactory family) and a gene for a novel
protein similar to 60S acidic ribosomal protein P2 (RPLP2),
complete sequence.
ACCESSION AL133267
VERSION AL133267.9 GI:10185396
KEYWORDS HTG: 7 transmembrane; olfactory receptor; ribosomal protein; RPLP2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 44788)
AUTHORS Williams, S.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 18, 2000 this sequence version replaced gi:9588470.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
validation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP3-408B20 is from the library RPCT-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RP3-408B20 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP1-31316 is at 44889 in this sequence.
The true right end of clone RP1-193B12 is at 100 in this sequence.
Location/Qualifiers
1..44788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-408B20"
/clone_lib="RPCT-3"
complement(1..100)
/note="match: STS: Em:HS193B12n"
222..535
/note="AluX repeat: matches 3..301 of consensus"
932..1870
/gene="dJ408B20.1"
932..1870
CDS
```

```
gb_hcg:AC074365 - 944.00 1296.69 6.0e-64 141268 | AC074365 Homo sapiens ch
em_hcg:hum:AL330860 - 944.00 1294.60 7.8e-64 178295 | AL330860 Human DNA seqeue
gb_hcg:AC091612 - 944.00 1294.49 7.9e-64 180657 | AC091612 Homo sapiens ch
gb_hcg:AL337039 + 944.00 1293.99 8.4e-64 190889 | AL337039 Homo sapiens ch
gb_pr:HS193612 + 941.00 1337.26 3.3e-66 936 | AJ302547 Homo sapiens 6M1-3
```

```

/gene="dj408B20.1"
/note="dj408B20.1 (novel 7 transmembrane receptor
(factory family) (hs6m1-33p) pseudogene)
match: CDNA: Em:M64392 Em:AF102516 Em:Y15524 Em:Y15525
Em:X89699 Em:AF102521 Em:AF042016 Em:AF044033 Em:AF102525
Em:AF044034 Em:AF044035 Em:AJ003145 Em:Y14442 Em:AF044037
Em:Y10529 Em:AF102529 Em:AF044038 Em:AF091565 Em:Y07557
Em:X89706 Em:AF042020 Em:M64391 Em:AF102532 Em:AF102533
Em:M64376 Em:AF102535 Em:M64377 Em:M64378 Em:AF102537
Em:AF091573 Em:AF091574 Em:AF091577 Em:AF091578
Em:AF102540 Em:M64385 Em:AF044053 Em:M64386 Em:M64387
Em:M64388 Em:AF091580 Em:DI2820
match: proteins: Tr:076000 Tr:076001 Tr:076002 Tr:035434
Sw:Q15062 Tr:Q95918 Tr:Q9T084 Tr:Q90806 Tr:Q9Y389
Tr:Q9WV13 Tr:Q95918 Tr:Q9T084 Tr:Q95006 Tr:Q9Y389
Tr:Q9Q222 Sw:Q95156 Sw:Q95157 Tr:Q9UGF6 Tr:Q9Y239
Tr:Q9UJ15 Tr:Q62007 Sw:P23275 Tr:Q95371 Tr:Q9UJ20
Tr:Q9Q218 Tr:Q9Q219 Tr:Q9PSJ3 Tr:P70526 Tr:Q9Q221"
/codon_start=1
/pseudo
evidence=not_experimental
complement(1518, 1963)
/note="match: GSS: Em:AQ546128"
2024, 2298
repeat_region
/note="Aluud repeat: matches 1, .280 of consensus"
2299, 2382
/note="42 copies 2 mer ag 79% conserved"
2559, 2650
repeat_region
/note="2 copies 46 mer 91% conserved"
complement(4227, 4576)
/misc_feature
/note="match: STS: Em:G31214"
4719, 5074
repeat_region
/note="L2 repeat: matches 2375, .2750 of consensus"
complement(4834, 5686)
/misc_feature
/note="match: GSS: Em:AQ748194"
5098, 5236
repeat_region
/note="L2 repeat: matches 1381, .1525 of consensus"
5707, 5843
repeat_region
/note="MIR repeat: matches 102, .240 of consensus"
6275, 6509
repeat_region
/note="L1M4 repeat: matches 5093, .5329 of consensus"
complement(6590, 6889)
/misc_feature
/note="match: GSS: Em:U94441"
6679, 6752
repeat_region
/note="MER76 repeat: matches 613, .685 of consensus"
6753, 6850
repeat_region
/note="L1P repeat: matches 1, .98 of consensus"
6820, 7010
repeat_region
/note="L1P repeat: matches 580, .776 of consensus"
7006, 7615
repeat_region
/note="L1P repeat: matches 900, .1509 of consensus"
7611, 9145
repeat_region
/note="L1P repeat: matches 3813, .5344 of consensus"
9751, 10411
repeat_region
/note="HERVL repeat: matches 3244, .3919 of consensus"
10555, 10663
repeat_region
/note="L1R16A repeat: matches 203, .307 of consensus"
10964, 11131
repeat_region
/note="L1R1H repeat: matches 1, .168 of consensus"
11705, 12141
repeat_region
/note="L1M4 repeat: matches 3898, .4367 of consensus"
13143, 13516
/misc_feature
/note="match: GSS: Em:AQ100659"
13668, 13803
repeat_region
/note="MIR repeat: matches 58, .199 of consensus"
complement(15225, 15506)
/misc_feature
/note="match: GSS: Em:AQ100750"
15627, 16228
/misc_feature
/note="match: GSS: Em:AQ544878"
16824, 17089
repeat_region
/note="Aluub repeat: matches 29, .311 of consensus"
17090, 17158
repeat_region
/note="Aluuf/FLAM repeat: matches 1, .81 of consensus"
17553, 17848
repeat_region
/note="AluX repeat: matches 3, .298 of consensus"
17789, 18240
/misc_feature
/note="match: STS: Em:G31213"
20021, 20249
repeat_region
/note="L1P1 repeat: matches 5904, .6155 of consensus"
20708, 21649
gene
/gene="dj408B20.2"
20708, 21649
CDS
/note="match: CDNA: Em:AF044034 Em:Y14442 Em:X89700
Em:Y07557 Em:AF042023 Em:AF091573 Em:AF091574 Em:X89670
Em:M64391 Em:AF102540 Em:AF010293 Em:M64388 Em:DI2820
Em:AF034896
match: ESTs: Em:N68399
match: proteins: Tr:Q43883 Tr:Q9Q221 Tr:Q9Z1K0 Tr:Q9Q220
Tr:Q83384 Tr:Q95409 Tr:Q9Q219 Sw:P23270 Tr:Q9Y387
Tr:Q9Q222 Tr:Q9Y299 Tr:Q62943 Tr:Q90807 Tr:Q90808
Tr:Q9WV13 Tr:Q63395 Tr:Q95918 Sw:P23266 Sw:P37069
Sw:Q95156 Tr:Q9UGF4 Tr:Q9WU86 Tr:Q60403 Tr:Q76000
Sw:P37070 Tr:Q9WU38 Tr:Q77758 Tr:Q9Q217 Tr:Q9R0K3
Tr:Q9R0K4 Tr:Q60412 Tr:Q9QY00 Tr:Q9WU36 Tr:Q9UJ20
Sw:P37072 Sw:P37071 Tr:Q9Y385"
/codon_start=1
evidence=not_experimental
/product="dj408B20.2 (novel 7 transmembrane receptor
(factory family) (hs6m1-32))"
/protein_id="CAC14158.1"
/db_xref="GI:10944516"
/translation="MNWVDSIIQEFILLGSDRPWLEFPLLVFLISYTVTIFGNLT
ILVSRIDTQLHTPMYFELTNLSLIDLCYTPOMLVNLCISIRKYSYSGCAOLF
IFLAGATEYLLAVMSFDRVAICRPLHYVIMHORCLQLAASWVTGFSNVTLS
TUTLQPLCDPYVIDHFEVCEPALKISCVETNENAEFLVSELPFILITLILISY
AFIVRAVRIOSASGRKARCTGCSHLIVLEFSTANASVVLQPPSPSKQCKMNVSL
FGIITAPMLNLIITLRNKEVKEGFKRLVAAYFLIKK"
21073, 21162
repeat_region
/note="L1R15 repeat: matches 127, .212 of consensus"
complement(21244, 21671)
/misc_feature
/note="match: GSS: Em:AQ077154"
21964, 21999
repeat_region
/note="18 copies 2 mer ca 100% conserved"
23289, 23944
repeat_region
/note="L1MB2 repeat: matches 5508, .6168 of consensus"
24195, 24291
repeat_region
/note="FLAM C repeat: matches 38, .136 of consensus"
24448, 24781
/misc_feature
/note="match: GSS: Em:B75689"
24537, 24857
repeat_region
/note="match: GSS: Em:AQ102724"
24739, 25051
repeat_region
/note="Alusg repeat: matches 1, .310 of consensus"
25407, 25944
repeat_region
/note="L1M4 repeat: matches 5206, .5763 of consensus"
25945, 26050
/misc_feature
/note="Alusg/x repeat: matches 195, .302 of consensus"
26055, 26632
misc_feature
/note="match: GSS: Em:AQ057601"
26059, 26126
repeat_region
/note="L1MB7 repeat: matches 6111, .6169 of consensus"
26164, 26220
repeat_region
/note="L1MB7 repeat: matches 6111, .6168 of consensus"
26221, 26391
repeat_region
/note="Aluub repeat: matches 137, .306 of consensus"
alignment_scores:
Quality: 1601.00 Length: 313
Ratio: 5.115 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-755-017-2 x AL133267
Align seg 1/1 to: AL133267 from: 1 to: 44788

```

1 MetAsnTrpValAlaAspSerIleIleGlnGluPheIleLeuGly 17  
 20708 ATGATTTGGGTAATGACAGCATCATACAGGATTATTCTGCTGGGTTT 20757  
 17 eSerAspArgProIleGluPheProLeuValValPheLeuIle 34  
 20758 CTCACATCGACCTTGGCTGGAGTTTCCACCTCTGGCTCTTGATTT 20807  
 34 eTyrThrValIlePheGlyAsnLeuThrIleIleLeuValSerArg 50  
 20808 CTTACACGTGACCATCTTGGCAATCGACCATATTCTAGTGCACGC 20857  
 51 LeuAspThrIleLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67  
 20858 CTGGACACCAACTTCATACCCCATGATTTTCTTACCAATCTATTC 20907  
 67 rLeuLeuAspLeuGlyCysTyrThrCysThrValProGlnMetLeuVal 84  
 20908 ACTCCTGGATCTTTGTACACCATGTACAGTCCACCAATGCTAGTAA 20957  
 84 snLeuGlySerIleArgIleValIleSerTyrArgGlyCysValAlaGln 100  
 20958 ATTATGACAGCATCAGAAAGTAATCATGTAATCGCTGTGTACCCAG 21007  
 101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117  
 21008 CTTTTCATATTTCTGGCTGGGGCTACTGATATCTTCTCTGCGCGCT 21057  
 117 MetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134  
 21058 CATGTCCTTTGATAGTTGTAGCATTTTGGCGCTCTCCATTTACTCAG 21107  
 134 aIleMetHisGlnArgLeuGlyLeuGlnLeuAlaIleAlaSerTrpVal 150  
 21108 TTATCATCTACACAGAGACTGCTCCAGTTGGCAGCTGCATCCYGGTT 21157  
 151 ThrGlyPheSerAsnSerValTyrPheSerThrIleThrLeuGlnLeuP 167  
 21158 ACTGTTTGTAGTACTGAGTGGTGTCTACCTGCTCTCCAGCTCCAGCTGC 21207  
 167 OlLeuGlyAspProTyrValIleAspHisPheLeuGlyGluValProAla 184  
 21208 ACTGTCGACCCCTATGTGATAGATCATCTTCTGTGTAAGTCCCTGCAC 21257  
 184 euleuGlyLeuSerCysValGluThrThrAlaAsnGlnAlaGluLeuPhe 200  
 21258 TGCTCAAGTTATCTTGTGTGAGACACAGCAAAATGAGGCTGAACTATTC 21307  
 201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIle 217  
 21308 CTTCGACGTAGCTCTTCATCTTAATACCCCTGACACATCATCTTATATC 21357  
 217 rTyrAlaPheIleValAlaArgAlaValLeuArgIleGlnSerAlaGluGly 234  
 21358 ATATGCTTTTATNTGTCCGACGAGTATGAGATACAGTCTGCTGAAGTC 21407  
 234 rGlnIleValAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 250  
 21408 GACAAAAAGCATTTGGGACATGCTGCCATCTAATTGTGCTCTCTT 21457  
 251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267  
 21458 TTTTATATGACGCGCTCTGTGTACCTGCAACCACTTGCCTCCAGCTC 21507  
 267 rLysAspGlnGlyMetValSerLeuPheTyrGlyIleIleAlaProM 284  
 21508 CAAGGACCAAGAAAGATGTTTCTCTCTATGCAATCATGACACCA 21557  
 284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300  
 21558 TGCTGAATCCCTTATATATACACTTAGAACCAAGAGGTAAGGAAGGC 21607

301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313  
 21608 TTTAAAGCTGTGTCAGAGAGCTCTTCTTAATCAAGAAA 21646

seq\_name: gb\_hlg:AC025941

seq\_documentation\_block: HNG 26-MAY-2000  
 LOCUS AC025941 176277 bp DNA  
 DEFINITION Homo sapiens chromosome 6 clone RP11-635011 map 6, WORKING DRAFT  
 SEQUENCE, 15 unordered pieces.  
 ACCESSION AC025941  
 VERSION AC025941.2 GI:7408057  
 KEYWORDS HNG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 (bases 1 to 176277)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 6, clone RP11-635011  
 Unpublished  
 2 (bases 1 to 176277)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,  
 Campiano,A., Castle,A., Chepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArillano,K., Dewar,K., Diaz,D.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hages,B., Heatford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lebecky,J.,  
 Levine,R., Leu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPeeters,R.,  
 Meldrum,J., Meneses,L., Milova,T., Miranda,C., Miengo,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisanu,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

#### TITLE

#### JOURNAL

#### COMMENT

Submitted (17-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 4, 2000 this sequence version replaced gi:7259782.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

#### Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

#### Project Information

Center project name: L8577  
 Center clone name: 635-O-11

#### Summary Statistics

Sequencing vector: MJ3; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 168925 bases at least Q40  
 Consensus quality: 172827 bases at least Q30  
 Consensus quality: 174151 bases at least Q20  
 Insert size: 176000; agarose-efp  
 Insert size: 174877; sum-of-ctrls  
 Quality coverage: 4.5 in Q20 bases; sum-of-ctrls  
 Quality coverage: 4.6 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is

```
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2997: contig of 2997 bp in length.
* 2998 3097: gap of 100 bp
* 3098 5606: contig of 2509 bp in length
* 5607 5706: gap of 100 bp
* 5707 9209: contig of 3503 bp in length
* 9210 9309: gap of 100 bp
* 9310 13614: contig of 4305 bp in length
* 13615 13714: gap of 100 bp
* 13715 20639: contig of 6925 bp in length
* 20640 20739: gap of 100 bp
* 20740 26545: contig of 5806 bp in length
* 26546 26645: gap of 100 bp
* 26646 37068: contig of 10423 bp in length
* 37069 37168: gap of 100 bp
* 37169 48083: contig of 10915 bp in length
* 48084 48183: gap of 100 bp
* 48184 59015: contig of 10832 bp in length
* 59016 59115: gap of 100 bp
* 59116 70420: contig of 11305 bp in length
* 70421 70520: gap of 100 bp
* 70521 80466: contig of 9946 bp in length
* 80467 80566: gap of 100 bp
* 80567 92325: contig of 11759 bp in length
* 92326 92425: gap of 100 bp
* 92426 110572: contig of 18147 bp in length
* 110573 110672: gap of 100 bp
* 110673 131898: contig of 21226 bp in length
* 131899 131998: gap of 100 bp
* 131999 176277: contig of 44279 bp in length.
*
* Location/Qualifiers
*   1..176277
*     /organism="Homo sapiens"
*     /db_xref="taxon:9606"
*     /chromosome="6"
*     /map="6"
*     /clone="RP11-635011"
*     /clone_lib="RPCI-11 Human Male BAC"
*     1..2997
*       /note="assembly_fragment"
*       3098..5606
*         /note="assembly_fragment"
*         clone_end:sp6
*         vector_side:left"
*         5707..9209
*           /note="assembly_fragment"
*           9310..13614
*             /note="assembly_fragment"
*             13715..20639
*               /note="assembly_fragment"
*               20740..26545
*                 /note="assembly_fragment"
*                 clone_end:r7
*                 vector_side:right"
*                 26646..37068
*                   /note="assembly_fragment"
*                   37169..48083
*                     /note="assembly_fragment"
*                     48184..59015
*                       /note="assembly_fragment"
*                       59116..70420
*                         /note="assembly_fragment"
*                         70521..80466
*                           /note="assembly_fragment"
*                           80567..92325
*                             /note="assembly_fragment"
*                             92426..110572
*                               /note="assembly_fragment"
*                               110673..131898
```

```
misc_feature      /note="assembly_fragment"
131999..176277
BASE COUNT      54493 a 33357 c 33254 g 53772 t 1401 others
ORIGIN
alignment_scores:
    Quality: 1601.00      Length: 313
    Ratio: 5.115          Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-09-755-017-2 x AC025941/rev
Align seg 1/1 to reverse of: AC025941 from: 1 to: 176277
1 MetAsTrrValAsnAspSerIleIleGlnGluPheIleLeuGlyPh 17
|||||
154174 ATGAATTGGGTAAATGACAGCATCATACAGAGATTATTCTGCGGTTT 154125
17 eSeAsPArGpRrOTRleuGluPheProLeuValAlaPheLeuIleS 34
|||||
154124 CTCAGATCGACCTTGGCTGAGTTTCACCTCTGTGGCTCTTGATT 154075
34 eTrrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
|||||
154074 CTACACTGAGACCATCTTGGCAATCTGACCATTTTCTAGTCACCC 154025
51 LeuAsPThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeuSe 67
|||||
154024 CTGGACACCAAACTTCATACCCCATGTATTTTCTTACCAATCTATC 153975
67 rIleuAsPleuCysTrrThrThrcysThrValProGlnMetLeuValA 84
|||||
153974 ACTCCTGGATCTTTGTTACACCATGTACAGTCCACAAATGCTAGTAA 153925
84 snLeuCysSerIleArgLysValIleSerTyrArgLysCysValAlaGln 100
|||||
153924 ATTATGACAGCATCAGCAAGATATCATGTTATTCGTGAGCCCGAG 153875
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGlyTrrLeuLeuAlaValA 117
|||||
153874 CTTTTCATATTTCTCGGCTTGGGGGCTACTGAATATCTTCTCGTGCCCT 153825
117 lMetSerPheAsPArgPheValAlaIleCysArgProLeuHisTyrSerV 134
|||||
153824 CATGTCCTTTGATAGGTTGTAGCTATTTGTGCGCTCTCCATTACTCAG 153775
134 alIleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaIleSerTrpVal 150
|||||
153774 TTATCATGACACGACGACACTCGCTCCACTTGGCAGCTCATCTCGGTT 153725
151 ThrGlyPheSerAsnSerValIrrPheSerThrLeuThrLeuGlnLeuPr 167
|||||
153724 ACTGGTTTAACTACACTGAGTGTGTCTTACCTGACACTCCAGCTGCGC 153675
167 oLeuCysAspProTyrValIleAspHisPheLeuCysGluValProAlaI 184
|||||
153674 ACTCTGTGACCCCTATGTGATGATCACTTCTGTGAGTCCCTGCGAC 153625
184 euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200
|||||
153624 TCCTCAAGTTATCTTGTGTGAGACACAGCAAAATGAGGCTCAACTATTC 153575
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSe 217
|||||
153574 CTTGTCAAGTACACTCTTCATCTAAATACCCCTGACACATCATCTTATATC 153525
217 rTrrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyA 234
|||||
153524 ATATGCTTTTATTGTCCAGACAGTATTGAGATACAGTCTGCTGAAAGTC 153475
```

234 rglulnalsalaphneglythrCysglySerHisleuleValSerleu 250  
|||||  
153474 GACAAAACATTGGGACATGGTCCCATCTATTTGTGCTCTTT 153425  
251 pheTySerThrAlaValSerValTyrLeuGlnProPseProSerse 267  
|||||  
153424 TTTTATAGACAGCGCTCTGTGTGACCAACCACTTCGCCAGCTC 153375  
267 rlysaapGlnGlyLysMetValSerleuPheTyrglylleleAlaPron 284  
|||||  
153374 CAAGGCCAGGAAAGATGGTTCTCTCTTATGGAACATTCACCA 153325  
284 etleuasnProleuileTyThrleuArgAsnlysglyVallysglygly 300  
|||||  
153324 TCGTGATTCCTCTTATATATACACTTACGACAGAGGTTAAAGGAGGC 153275  
301 pheLysArgLeuValAlaArgValPheleuileLysLys 313  
|||||  
153274 TTTTAAAGGTGGTTCAGAGAGTCTCTTAATCAAGAAA 153236  
seq\_name: gb\_HSA302593  
seq\_documentation\_block:  
LOCUS HSA302593 1074 bp DNA PRI 06-JAN-2001  
DEFINITION Homo sapiens 6M1-10\*02 gene for olfactory receptor, cell 11ne  
KR3598.  
ACCESSION AJ302593  
VERSION AJ302593.1 GI:12054410  
KEYWORDS 6M1-10\*02 gene; olfactory receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1074)  
AUTHORS Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,  
Uchanska-Ziegler,B., Volz,A., Younger,R., and Beck,S.  
TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(in) Kasahara,M.(Ed.):  
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
FUNCTION: 110-130;  
Springer-Verlag, Tokyo, Japan (2000)  
REFERENCE 2 (bases 1 to 1074)  
AUTHORS Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,  
Younger,R.M. and Beck,S.  
TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
Unpublished  
3 (bases 1 to 1074)  
AUTHORS Younger,R.M., Amadou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,  
Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,  
Volz,A., Ziegler,A. and Beck,S.  
TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
in Human and Mouse  
Unpublished  
4 (bases 1 to 1074)  
AUTHORS Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.  
and Ziegler,A.  
TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes  
Unpublished  
5 (bases 1 to 1074)  
AUTHORS Ehlers,A.  
TITLE Direct Submission  
JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
REFERENCE Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
JOURNAL GERMANY  
FEATURES  
SOURCE location/Qualifiers  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="KR3598"  
gene

/gene="6M1-10\*02"  
1..1074  
/gene="6M1-10\*02"  
/codon\_start=1  
/product="olfactory receptor"  
/protein\_id="CAC20513.1"  
/db\_xref="GI:12054411"  
/translation="MNNVNRKSVPOEFLLVSDPQWLEIPFVWFLESLITIFGNLT  
IILVSHDFKLIHTPWFPLSLJSLDLCYTSYVPMVLVNCIRKVISIGGVAAQLP  
IFLALGSTECILLAVMCFDFEVALICRPLVSIIMHORLCEQLAASMTSGFSNSVLDS  
TWLTKMPLCGHKEVDHFECEVPALIKSCVDTTNEAELEFISVFLIPYLLITLISY  
AFIVQAVLRIOSAGRRKAEFCGSHLIVSLFYGTAISMVLOPPSPSKRGKMWSL  
FCGLIAPMLNPLIYTLINKREKFAFKRLVANSLSLNGEIRNMQMSFAKDIYLYLTNPF  
SASCPPIVYITTECNLPORKEP"  
BASE COUNT 277 a 248 c 203 g 346 t  
ORIGIN  
Alignment\_scores:  
Quality: 1323.00 Length: 314  
Ratio: 4.531 Gaps: 1  
Percent Similarity: 92.994 Percent Identity: 82.803  
alignment\_block:  
US-09-755-017-2 x HSA302593  
Align seg 1/1 to: HSA302593 from: 1 to: 1074  
1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGlyPh 17  
|||||  
1 ATGAATTGGGTAATGAAGACTGTCGCCAGAGACTTCTGTACTTTT 50  
17 eSerAspArgProTrpLeuGluPheProleuValAlaPheLeuLys 34  
|||||  
51 CTCAGATCAACCATGGCTGAGATTCACCACTTGGATGTTCTGTTT 100  
34 eTyThrValThrIlePheGlyAsnLeuThrIleLeuValSerArg 50  
|||||  
101 CCTATATCTTGACAATCTTGGCAATCTGCACATTAATCTGTGCACAT 150  
51 LeuAspThrLysLeuHisThrPrometTyrrPhePheLeuThrAsnLeu 67  
|||||  
151 GTGGATTTCACAACTCCACACCCCTATGACTTTTCTTAGACATCTCTC 200  
67 rleuLeuAspLeuCysTyrrThrCysThrValProGlnMetLeuVal 84  
|||||  
201 ACTCTGGACCTTGTCTATACCAAGTACAGTCCACAATCTCGTAA 250  
84 snleuGlySerIleArgLysValIleSerTyrrGlyCysValAlaGln 100  
|||||  
251 ACATATGCAACACCAAGGAAAGTATCATGTTGTGGCTGTGGCCAG 300  
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGlyTyrLeuLeuValAla 117  
|||||  
301 CTTTATATTTCTCGGCTTGGCTTGGCTTCCACAGAAATGCTTCTCGGCCGT 350  
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrrSer 134  
|||||  
351 CATGCTCTTGTATAGTTGTGTACTTATGTCGCGCTCCCATTTACTCAA 400  
134 alileMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaAsertTrpVal 150  
|||||  
401 TTATATGACACGAGAGCTGTGCTTCAGTTGGACCTGCATCTGATTT 450  
151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuP 167  
|||||  
451 AGTGGCTTTAGCATTCAGATTACAGTCCACCTGCAGACACTTAAGATGCC 500  
167 oleuGlyAspProTyrrValIleAspHisPheLeuGlyValProAla 184  
|||||  
501 ACTGTGTGTACAAAGAGATGATCATCTTCTGTGAAGTCCCGCTC 550  
184 euleuGlyLeuSerCysValGlyThrThrAlaAsnGlnAlaGluLeuPhe 200

```

|||||
551 TGCCAGAGTGTCTCTGTGACACACACAAATGAGCGTGAACATATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuLeuIle 217
601 TTTCAATGCTGTCTATTCCTTTAATACCCGTGACCTCATCTTATATC 650
217 TTTTATAPheIleValArgAlaValLeuArgIleGlnSerAlaGluGly 234
651 GTATGCTTTATTTGTCACAGCAGTGTGAGATCCACTGCTGTAAGCTC 700
234 rglLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeu 250
701 GACGAAGCGATTTGGACATGTGGCTCCCATCAATTTGCTGCTACTT 750
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267
751 TTTTATGTTACAGCTATCTCCATGCTCAACACCTTCACCCAGCTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPhe 284
801 CAAGACGCGGGAAGATGGTTCTCTCTCTGTGGAATCATTCGACCCA 850
284 etLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGly 300
851 TGTGGAATCCCTTATATATACCTTAGCAACAAAGAGTAAGAGGACC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLys 313
901 TTTTAAAGGTTGGTTGCCAAGAGCTTCTTATATCAGAAA 940
seq_name: gb_pr: HSA302584
seq_documentation_block:
LOCUS HSA302584 1074 bp DNA PRI 06-JAN-2001
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line
BM28.7.
ACCESSION AJ302584 GI:12054392
VERSION 6M1-10*01 gene; olfactory receptor.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Younger,R.M. and Beck,S.
1 (bases 1 to 1074)
UCHANSKA-ZIEGLER,B., VOLZ,A., YOUNGER,R. and BECK,S.
Polymorphic olfactory receptor genes and HLA loci constitute
extended haplotypes
(in) Kasahara,M. (Ed.):
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
FUNCTION: 110-130:
Springer-Verlag, Tokyo, Japan (2000)
REFERENCE
AUTHORS Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Younger,R.M. and Beck,S.
1 (bases 1 to 1074)
POLYMORPHISMS IN OLFACTORY RECEPTOR GENES: A CAUTIONARY NOTE
Unpublished
3 (bases 1 to 1074)
YOUNGER,R.M., AMADOU,C., BETHEL,G., EHLERS,A., FISCHER LINDAHL,K.,
FORBES,S.A., HORTON,R., MILNE,S., MUNGALL,A.J., TROWSDALE,J.,
VOLZ,A., ZIEGLER,A. and BECK,S.
Characterisation of clustered MHC-linked Olfactory Receptor Genes
in Human and Mouse
Unpublished
4 (bases 1 to 1074)
EHLERS,A., BECK,S., FORBES,S.A., TROWSDALE,J., VOLZ,A., YOUNGER,R.
and ZIEGLER,A.
MHC-linked olfactory receptor loci exhibit polymorphism and
contribute to extended HLA/OR-haplotypes
Unpublished
5 (bases 1 to 1074)
EHLERS,A.
JOURNAL
REFERENCE
AUTHORS

```

```

TITLE Direct Submission
JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,
GERMANY
FEATURES
source Location/Qualifiers
1..1074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/cell_line="BM28.7"
1..1074
/gene="6M1-10*01"
1..1074
/gene="6M1-10*01"
/codon_start=1
/product="olfactory receptor"
/protein_id="CAC20504.1"
/db_xref="GI:12054393"
/translocation="MMWVKSEVPOEFLIVSDQPMLEIPPFVMEFLFSLITFENLT
ILVSHDFEKLHTPEYFLSNLSLDLCYTSVDMLEIPNLCNTRKVISYGVCAOLF
IEIALGSTRECLILVCMCPDRVALICRPLHYSITIMHQRCPOLAAASVTSQSVNLOF
TWTLKMPICGRKENDHFCFCEPALIKLSVDTTANBAELFISVLFLIPYTLILIS
AFIVQAVLRIOASBQRKAFGTCGSHLIVSLFTGTAISMTLOPSPSKGRKMWSL
FCGIAPMLNPLIYTLRNKREKAEFRILVASSLLNOETIRNMOMISFAKDTLVLYLTNF
SASCPIFVITTEYCNLPORKEP"
BASE COUNT 278 a 248 c 202 g 346 t
ORIGIN
alignment_scores:
Quality: 1319.00 Length: 314
Ratio: 4.517 Gaps: 1
Percent Similarity: 92.994 Percent Identity: 82.484
alignment_block:
US-09-755-017-2 x HSA302584 ..
Align seg 1/1 to: HSA302584 from: 1 to: 1074
1 MetAsnTrpAlaInsnspsSerIleIleGlnGluPheIleLeuGlyPh 17
|||||
1 ATGAATTGGGGTAATAAGAGTGTCCACAGAGACTTATCTGTAGATT 50
17 eSerAspArgProTrpLeuGlnPheProLeuValAlaPheLeuIle 34
|||||
51 CTGAGATGACACCATGGCTAGAGATTCACCCCTTGGATGATTTCTGTTT 100
34 eTyrThrValTrpIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
|||||
101 CTTATATCTTGACAAATCTTGGCAATCTGCACATAATTTCTGTGTCACAT 150
51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67
|||||
151 GTGGATTTCAACTCCACACCCCTATGTACTTTTCTTTCGACAAATCTCTC 200
67 rLeuLeuAspLeuCysTyrThrThrCysTrpValProGlnMetLeuVal 84
|||||
201 ACTCTGTGACCTTGGCATATCCACAGTACAGTTCACAAATGCTGTAA 250
84 snLeuCysSerIleArgLysValIleSerTyrArgGlyCysValAlaGln 100
|||||
251 ACATATGACACACACAGGAAGATGATGATGATGCTGTGCGCCAG 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAla 117
|||||
301 CTTTTCATTTTCTGGCTGGGGGTCCACAGAAATGTCTTCTCGGCCGT 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 134
|||||
351 CATGTGCTTGTATAGGTTTGTACATTTTGTGGCTTCATTTACTCAA 400
134 alIleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaSerTrpVal 150
|||||

```



401 TTATCATGCCACGAGCTCTGCTTCAGTTCAGACCTCATCTCGATT 450  
 151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeu 167  
 451 AATGGCTTTAGCAATTCATGATTTACAGTCACCTGACACTTAAGATGCC 500  
 167 OleuGlyAspProTyrValIleAspHisPheLeuGlyValProAla 184  
 501 ACTGTGTGTCACAAAGAGTGCATCTCTTCTGTGTGAAGTCCCTGCTC 550  
 184 euleuGlyLeuSerGlyValGluThrThrAlaAsnGluAlaGluLeu 200  
 551 TGCTCAAGTTGTCTGTGTGTGACACAGCAATGAGAGCTGACATATTC 600  
 201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeu 217  
 601 TTCATCAGTGTCTATTCCTTAATACCCGACACTCATCTATATATC 650  
 217 TyrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGly 234  
 651 GATGCTTTATTTGTCACAGCAGTGTGTGAATCCAGTCTGCTGAAGTTC 700  
 234 rglGlnLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 250  
 701 AACGAAAGCATTTGGACATGTGGCTCCCATCTATTTGTGTGTCATCT 750  
 251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267  
 751 TTTTATGAGTACAGTATCTCATGACGCAACCACTTCACCCAGCTC 800  
 267 rTysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284  
 801 CAAGACCGGGAGAGATGTTCTCTCTCTGTGTGAATCATTCACCA 850  
 284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGlu 300  
 851 TGCTGATCCCTTATATATACACTTAGAACAAGAGTAAGAGAGCC 900  
 301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313  
 901 TTTTAAAGGTGTGTCAAAGAGTCTTCTTAATCAAGAA 940  
 seq\_name: gb\_HSA302585  
 seq\_documentation\_block:  
 LOCUS HSA302585 1074 bp DNA PRI 06-JAN-2001  
 DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell 11ne  
 BM19.7.  
 ACCESSION AJ302585  
 VERSION AJ302585.1 GI:12054394  
 KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1074)  
 AUTHORS Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,  
 Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.  
 TITLE Polymorphic olfactory receptor genes and HLA loci constitute  
 extended haplotypes  
 JOURNAL (in) Kasahara,M.(Ed.):  
 MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND  
 FUNCTION: 110-130;  
 SPRINGER-Verlag, Tokyo, Japan (2000)  
 REFERENCE 2 (bases 1 to 1074)  
 AUTHORS Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,  
 Younger,R.M. and Beck,S.  
 TITLE Polymorphisms in olfactory receptor genes: a cautionary note  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1074)  
 AUTHORS Younger,R.M., Anadou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,  
 Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,  
 Volz,A., Ziegler,A. and Beck,S.

TITLE Characterisation of clustered MHC-linked Olfactory Receptor Genes  
 in Human and Mouse  
 JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 1074)  
 AUTHORS Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.  
 and Ziegler,A.  
 TITLE MHC-linked olfactory receptor loci exhibit polymorphism and  
 contribute to extended HLA/OR-haplotypes  
 JOURNAL Unpublished  
 REFERENCE 5 (bases 1 to 1074)  
 AUTHORS Ehlers,A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
 Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
 GERMANY  
 FEATURES  
 source Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="BM19.7"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /codon\_start=1  
 /product="Olfactory receptor"  
 /protein\_id="CAC20505.1"  
 /db\_xref="GI:12054395"  
 /translation="MMWVNSVPOEFLILVSDPMLPEPFWMLRSLYILITFNGLT  
 ILVSHVDFKLLHPMPYFLSNLSLDLCYTTSTYPMOLVINCNRKVIYSGCYAOLF  
 FLIALGSTECLILAVMCFDPVATCRPLHLSITHOGLCOLAASWISPSNVLDS  
 TWTLKMPLOCGRKVEDHFCERPAALKLSCDTTANEDLEFISVLLIPVTLILSY  
 AFIVQAVLRIOASAGQRARFETGSHLIVSLFETGTAISWYLOPSPSSKORGMVSL  
 FCGIAPMLNPLIYLRNKEVKAERKRLVAKSLNDEIRNMQMSIFAKVDVLYLTNMF  
 SASCPDIVITLENCNLPORKEP"  
 BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 1319.00 Length: 314  
 Ratio: 4.517 Gaps: 1  
 Percent Similarity: 92.994 Percent Identity: 82.484  
 alignment\_block:  
 :US-09-755-017-2 x HSA302585 ..  
 Align seg 1/1 to: HSA302585 from: 1 to: 1074  
 1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGlyPh 17  
 1 ATGAATTTGGTAATAATAGAGTGTCCACAGAGATTCATCTGTAGTTT 50  
 17 eSerAspArgProTrpLeuGluPheProLeuLeuValAlaPheLeuLys 34  
 51 CTCAGATCAACCATGGCTAGAGATTCACCTTGTGAGATGTTCTGTTT 100  
 34 eTyrThrValThrIlePheGlyAsnLeuThrIleLeuValSerArg 50  
 101 CCTATATCTTGACAAATCTTTGGCAATCTGACATAATTTCTGTGTACAT 150  
 51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67  
 151 GTGATTTCAAACTCCACACCCCTATGACTTTTCTTAGCAATCTCTC 200  
 67 rLeuLeuAspLeuGlyTyrThrThrCysThrValProGlnMetLeuVal 84  
 201 ACTCCCTGGACCTTGTCTATACCAAGTACAGTACACAAATGCTGTGTA 250  
 84 snLeuGlySerIleArgLysValIleSerTyrArgLysValAlaGln 100  
 251 ACATATGCAACACGAGAAAGTATCATGTTATGTGTGTGTGTGTGTG 300

```

101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluThrLeuLeuAla 117
    |||||||
301 CTTTCAATTTTCCTGGCCTGGGCTGCACAGAAATGCTTCTCGGGCGGT 350
    |||||||
117 lmetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 134
    |||||||
351 CATGTGGTGGATAGGTTTGTACTATTTGTCCGCCCTCCATTAACA 400
    |||||||
134 alIleMetHISGlnArgLeuCysLeuGlnLeuAlaAlaIleSerTrpVal 150
    |||||||
401 TTATCATGACCCAGAGGCTCTGCTCCAGTTGGAGCTGCATCTCGATT 450
    |||||||
151 ThrGlyPheSerAsnSerValIrrPleSerThrLeuThrLeuGlnLeu 167
    |||||||
451 AGTGGCTTATGCAATTCAGTATTACAGTCCACCTGGACACTTAAGATGCC 500
    |||||||
167 oLeuCysAspProTyrValIleAspHisPheLeuCysGluValProAla 184
    |||||||
501 ACTGTGTGTGCACAAAGAGTGGATCATCTTCTGTGAAGTCCCTGCTC 550
    |||||||
184 euleuLysLeuSerCysValGluThrThraLAsnGlnAlaGluLeuPhe 200
    |||||||
551 TGCCTCAAGTTGCTCTCTGTGTGGACACACACCAATGAGCTGAACATATTC 600
    |||||||
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuLeuLeuIle 217
    |||||||
601 TTCATCAGTGTGCTATTCCTTAATACCGGTCACACTCATCTTAATATC 650
    |||||||
217 rTyraIlePheIleValAlaGlnValLeuArgIleGlnSerAlaGluGly 234
    |||||||
651 GTATGCTTTTATTTGTCACACAGCTGTGAAATCCAGTGTGTAAGGTC 700
    |||||||
234 rGlnLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeu 250
    |||||||
701 AACGAAAGGATTTGGGACATGTGCTCCCATCTAATGTGGTGCATCTT 750
    |||||||
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267
    |||||||
751 TTTTATGATGACAGCTATCTCATGATGACAGCAACCTCACACAGCTC 800
    |||||||
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284
    |||||||
801 CAAAGCCGGGGAAGATGGTTCCTCTCTGTGAAATCATTTGCAACCCA 850
    |||||||
284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
    |||||||
851 TGTGTGATCCCTTATATATACCTTAGCAACAAGGTAAGGAAGC 900
    |||||||
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
    |||||||
901 TTTAAAGGTTGGTGCAAAGAGTCTTAAATCAAGAAA 940
    |||||||
seq_name: gb_pr: HSA302586
seq_documentation_block:
LOCUS      HSA302586      1074 bp      DNA      PRI      06-JAN-2001
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line LG2.
ACCESSION  AJ302586
VERSION     AJ302586.1 GI:12054396
KEYWORDS   6M1-10*01 gene; olfactory receptor.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1074)
AUTHORS     Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J.,
            Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.
TITLE       Polymorphic olfactory receptor genes and HLA loci constitute
            extended haplotypes
JOURNAL     (in) Kasahara,M. (Ed.):
            MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
            FUNCTION: 110-130;

```

```

REFERENCE   2 (bases 1 to 1074)
AUTHORS     Springer-Verlag, Tokyo, Japan (2000)
TITLE       Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
            Younger,R.M. and Beck,S.
JOURNAL     Polymorphisms in olfactory receptor genes: a cautionary note
REFERENCE   3 (bases 1 to 1074)
AUTHORS     Younger,R.M., Amadou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,
            Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,
            Volz,A., Ziegler,A. and Beck,S.
TITLE       Characterisation of clustered MHC-linked Olfactory Receptor Genes
            in Human and Mouse
JOURNAL     Unpublished
AUTHORS     Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.
            and Ziegler,A.
TITLE       MHC-linked olfactory receptor loci exhibit polymorphism and
            contribute to extended HLA/OR-haplotypes
JOURNAL     Unpublished
AUTHORS     Ehlers,A.
TITLE       Direct Submission
JOURNAL     Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,
            Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,
            GERMANY
FEATURES
SOURCE      Location/Qualifiers
            1..1074
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /cell_line="LG2"
            1..1074
            /gene="6M1-10*01"
            1..1074
            /gene="6M1-10*01"
            /product="Olfactory receptor"
            /protein_id="CAC20506.1"
            /db_xref="GI:12054397"
            /translation="MNNVKNKSPDFLLVRSDDPMLEIFPPWFLLSYIITITGNLT
            IILVSHDFKLTPTWTFPLSLSLDLCYTSYTPQMLVNCNKRKVISYGGVQALF
            IFLVSGTECLLLAVMCFDRFVAICRPLHSIIIMHORICQDLAASWISGFSNVLGS
            TWTLKMPICGHEVDHFECEVPALKLSVDPTANEAELFISVFLIIPVTLILISY
            AFIVQAVLRISAGORAKFCGSHLIVSLFPGTAISWLOPPSPSSKDRGKWSL
            FCGIAPMLNPILYITLRNKEKFAKRLVAKSLNGLNQRNNQMSFPAKDYLVTLTNP
            SASPIRVITLLENCLNLPQRKEP"
BASE COUNT  278 a      248 c      202 g      346 t
ORIGIN
alignment_scores:
Quality: 1319.00      Length: 314
Ratio: 4.517          Gaps: 1
Percent Similarity: 92.994      Percent Identity: 82.484
alignement_block:
US-09-755-017-2 x HSA302586 ..
Align seg 1/1 to: HSA302586 from: 1 to: 1074
1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGlyPh 17
|||||
1 ATGAATTTGGTAAATAGAGTGTCCACAGAGGATTCATTCTGTACTTT 50
|||||
17 eSerAspArgProTyrPleuGluPheProLeuValAlaPheLeuLys 34
|||||
51 CTCAGATCAACCATGCTAGAGATTCACCTTTGGAGTGTCTGCTTTT 100
|||||
34 eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
|||||
101 CCTATATCTTGACAAATCTTTGGCAATCTGACATATATTTCTGTGCACAT 150
|||||
51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeuSe 67

```



1 ATGAATGGGGTAAATAGAAGGTGTCSSACAGAGATTCAATCTGTAAGTTT 50

17 eSerAspArgProThrLeuGlnPheProLeuValValPheLeuIleS 34

51 STCAATCATCAACCATGGCTGAGATTCACCCCTTTGGATGTTCTGGTTT 100

34 eTyrThrValThrIlePheGlyAspLeuThrIleLeuValSerArg 50

101 CSTATATCTGGACAATCTTTGGCAATCTACAAATTAATCTTGTGCATAT 150

51 LeuAspThrIysLeuHisThrProMetCysTrpPheLeuThrIyrAsnLeuSe 67

151 GTGATTTCAAACTCCACAGCCSCTATGACTTTTCTTACAGCATCTCTC 200

67 rLeuLeuAspIleCysTyrThrThrCysThrValProGlnMetLeuValA 84

201 ACTCTGTGACCCCTTCTATACCAACAAGTACAGTTCCACAATCTGGTAA 250

84 snLeuCysSerIleArgIysValIleSerTyrArgGlyIysValAlaGln 100

251 ACATATGCAACACCCAGGAAGTAATCAGTATAGTGCTGTGTGGCCAG 300

101 LeuPheIlePheLeuAlaLeuGlyAlaThrGlnTyrLeuLeuLeuAlaVal 117

301 CTATTCAATTTCTCGGCCCTGGGTTCCACAGAAATGCTTCTCTGGCCCT 350

117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134

351 CATGTGCTTTGTATAGCTTTGTAGCTAATTTGTGGCCCTCTCCATTTACTCA 400

134 alIleMetHisGlnArgIleuCysLeuGlnLeuAlaAlaAlaSerTyrVal 150

401 TTATCATCATCCACAGAGCCCTCTCTCCATCTGGCAGCTCATCTGTGATT 450

151 ThrGlyPheSerAsnSerValTyrPheSerThrLeuThrLeuGlnIlePur 167

451 AGTGGCTTTAGCAATTCAAGTATACAGTCCACCTGGACACCTTAAGAATGCC 500

167 oLeuCysAspProTyrValIleAspHisPheLeuCysGlyValProAlaL 184

501 ACTGTGGCTGACCAAGAAGAGTGCATCACTTTCTTGTGAAGTCCCTGCTC 550

184 euLeuIlyLeuSerCysValGlnThrThrAlaAsnGlnAlaGluLeuPhe 200

551 TGCCTCAATGTGCTCTGTGTGCACACAGCAAAATGAGGCTGAACATTTC 600

201 LeuValSerGluLeuPheHisIleuLeuPheProLeuThrLeuIleLeuIleSe 217

601 TTTCATCACTGTGCTATCTCTCTTAAATACCCGAGACACTATCTTATTATC 650

217 tTyrAlaPheIleValAlaGlyAlaValLeuArgIleGlnSerAlaGluGlyA 234

651 GTATGCTTTTATTGTCCCAAGCAGTGTGAGATTCAGTCTGGCGAAGGTC 700

234 rGlnGlnIlyAlaPheGlyIlyThrCysGlySerHisLeuIleValValSerLeu 250

701 AACCAAAAGGCAATTTGGGACATGTGGCTCCCACTAATGTGGTGCATCT 750

251 PheTyrSerThrAlaValSerValTyrLeuGlnIleProProSerProSerSe 267

751 TTTTATATGTTACAGCTATCTCCATGACCTGGCAACACCTTCAACCCAGCTC 800

267 rLysAspGlnGlyIlyMetValSerLeuPheTyrGlyIleIleAlaProM 284

801 CAAGAGACCGGGGAAGAAGATGGTTTCTCTTCCTGTGGAAATCATTTGCCCA 850

284 eIleuAsnProLeuIleTyrThrLeuArgAsnIlyGlyValIlyGlnGly 300

851 TGTGCAATCCCTTATATATATACACTTAGACAAAGAGGTAAAGAACCC 900

301 PheIysArgLeuValAlaLArgValPheLeuIleIlyLys 313

901 TTTTAAAGGTGGTTGGCAAAAGAGTCTTCTTAATCAAGAA 940

seq\_name: gb\_pr:HSA302588

seq\_documentation\_block:

LOCUS	HSA302588	1074 bp	DNA	PRI	06-JAN-2001
DEFINITION	Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line WT51.				

KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
SOURCE human.

ORGANISM

REFERENCE  
1 (bases 1 to 1074)  
Eukaryota: Metazoa: Chordata: Cranialta: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.

## AUTHORS

Uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(in) Kasahara, M. (Ed.):  
JOURNAL

JOURNAL

Springer-Verlag, Tokyo, Japan (2000)

## AUTHORS

**TITLE** Younger, R.M. and Beck, S.  
Polymorphisms in olfactory receptor genes: a cautionary note

JOURNAL  
REFERENCE

REFERENCE	AUTHORS	TITLE
3 (pages 18 to 104)	Younger, R.M., Amadou, C., Bethel, G., Ehlers, A., Fischer Lindahl, K., Forbes, S.A., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J., Volz, A., Ziegler, A. and Beek, S.	Characterisation of clustered MHC-linked Olfactory Receptor Genes

TITLE

UNPUBLISHED  
JOURNAL  
REFERENCE  
AUTHORS  
4 (bases 1 to 1074)  
Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R.

ENTRUSTED

TITLE	JOURNAL REFERENCE
mhc-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes	Unpublished 5 (bases 1 to 1074)

## AUTHORS

**TITLE** Direct Submission  
**JOURNAL** Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY

## FEATURES

Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/cell_line="WT51"
1. .1074
gene

```

gene

```

1.1074
/gene="6M1-10*01"
CDS

```

CDS

```

/gene="6M1-10*01"
/codon_start=1
/product="olfactory receptor"
/protein_id="CAC20508.1"

```

translation="MMWVNSVPQOEILILVFSDDPMLPEIPFVMELEFSLITLFIENGLT  
/ILVSHVDPLKHTPMYFSLMLSLDLCYTVSDPQMLNCFNTRVIVSGCVAOLF  
FIALGSTECCILAVMCFDEPRVAIKRIPLHYSIMHRCOLQOLAASMSISGNSNVYS  
TWLTPMPLCGHKHPVDFHCFCEPAALIKISCDTNTAAEAFIEIVSFLIPVPIITISY

BASE COUNT	278 a	248 c	202 g	346 t
ORIGIN				
				FCGIIAPMLNPLIYTLRNKEVKEAEAKRR
				SASCPIFVITIENYCNLPQRKFP <sup>a</sup>

ORIGIN

```
alignment_scores:
    Quality: 1319.00
    Length: 314
```

Length:

314

Ratio: 4.517 Gaps: 1  
Percent Similarity: 92.994 Percent Identity: 82.484

alignment block:  
US-09-755-017-2 x HSA302588 ..

Align seg 1/1 to: HSA302588 from: 1 to: 1074

```

1 MetAspTrpValAsnAspSerIleIleGlnGluPheIleLeuGluLys 17
1 ATGAATGGGTAATAGAGTGTCCACAGAGAGTGTCTTGTAGTATTT 50
17 eSerAspArgProTrpLeuGluPheProLeuValValPheLeuIles 34
51 CTCAGATCAACACATGGCTAGAGATTCACCCCTTGTGATGTTCTTT 100
34 eTrpThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
101 CCTAATATCTTGACAAATCTTGGCAATCTGACATATATCTTGTGCACAT 150
51 LeuAspThrLysLeuHisThrProMetLysTrpPheLeuThrAsnLeu 67
151 GGGATTTCAAACTCCACACCCCTATGTACTTTTCTTTCAGCAATCTCTC 200
67 rLeuLeuAspLeuCysTrpThrThrCysThrValProGlnMetLeuVal 84
201 ACTCCGGACCTTGTGCTATACCAAGTACAGTCCACAATGCTGTGTA 250
84 snLeuCysSerLeuArgLysValIleSerTyrArgGlyCysValAlaGln 100
251 ACATATGCAACACACAGAAAGTAATCAGTATGCTGTGCTGTGGCCAG 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuVal 117
301 CTTTTCATTTCTCGCTGGCTTGGCTCCACAGAAATGCTTCTCTGCGCGT 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
351 CATGTGCTTTGATAGTGTGTAGCTATTTGTGCGCTCTCCANTACTCA 400
134 aIleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaAspTyrVal 150
401 TTATCATGACACACAGAGCTGTGCTTCACAGTGCACCTCATCTTGAT 450
151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeu 167
451 AGTGGCTTTAGCAATTCAGTATATACAGTCCACCTGAGCACTTAAGATGCC 500
167 oLeuCysAspProTyrValIleAspHisPheLeuCysGluValProAla 184
501 ACTGTGTGTCACAAAGAGTGTATCTTCTGTGAAGTCCCTGCTC 550
184 euleuLeuLysSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200
551 TCCTCAAGTGTCTGTGTGACACAAAGCAATGAGGCTGAACATATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIle 217
601 TTCATAGAGTGTATCTCTTCAATATCCGAGTACACTCATCTTATATC 650
217 rTyrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGly 234
651 GATGCTTTTATTTGTCACAGAGTGTGAAGATCCAGTCTGCGAAGTTC 700
234 rGlnIleLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 250
701 AACGAAAGCATTTGGACATGTGGCTCCCATCTAATTTGGTGCATCT 750
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267
751 TTTTATGACAGACTATCTCATGATGACGACCAACCTTCACCCAGCTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284

```

```

|||||.....
801 CAAGACCGGGGAAAGATGTTCTCTCTGTGTGAATCATTCACACCA 850
284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
851 TGCTGAATCCCTTATATATACACTTAGAACAAGAGTAAAGGAAGCC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
901 TTTAAGGTGTTGTTCCAAAGAGTCTTCTTATATCAAGAA 940
seq_name: gb_pr:HSA302589
seq_documentation_block:
LOCUS HSA302589 1074 bp DNA PRI 06-JAN-2001
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line SA.
ACCESSION AJ302589
VERSION AJ302589.1 GI:12054402
KEYWORDS 6M1-10*01 gene; olfactory receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J.,
Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.
Polymorphic olfactory receptor genes and HLA loci constitute
extended haplotypes
(in) Kasahara,M. (Ed.):
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
FUNCTION: 110-130;
Springer-Verlag, Tokyo, Japan (2000)
2 (bases 1 to 1074)
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Younger,R.M. and Beck,S.
Polymorphisms in olfactory receptor genes: a cautionary note
Unpublished
3 (bases 1 to 1074)
Younger,R.M., Amadou,C., Bethel,G., Ehlers,A., Fischer,Lindahl,K.,
Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,
Volz,A., Ziegler,A. and Beck,S.
Characterisation of clustered MHC-linked olfactory Receptor Genes
in Human and Mouse
Unpublished
4 (bases 1 to 1074)
Ehlers,A., Beck,S., Forbes,S.A., Trowsdale,J., Volz,A., Younger,R.
and Ziegler,A.
MHC-linked olfactory receptor loci exhibit polymorphism and
contribute to extended HLA/OR-haplotypes
Unpublished
5 (bases 1 to 1074)
Ehlers,A.
Direct Submission
Submitted (06-NOV-2000) Ehlers A., Institut fuer Immungenetik,
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,
GERMANY
FEATURES
Source
Location/Qualifiers
1..1074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/cell_line="SA"
1..1074
/gene="6M1-10*01"
1..1074
/gene="6M1-10*01"
/codon_start=1
/product="Olfactory receptor"
/protein_id="CAC20509.1"
/db_xref="GI:12054403"
/translation="MNWVNRKSVPOEFILVSDPWFLEIPFVFLFSYLLTFGNLT
IIVSHVDFKLTPTMVFPLSNLSLIDCYTSTVPMQVNTCRKRVISYSGCVAQLF
IFLAGSTECILLVWCPDRVAICRPLHYSIIMHQRLCPOLAAASWISGFSNVLAGS

```

TWTLKMP LCGHEVDHFECEVPA LKLSVDITNENAE LFFISVLFIPTVLLISY  
AFIYQAVLRISASBQKAFGTCGSHLIVSLFEGTASMLQPPSPSKRGMVSL  
FCGIAPLNPILITLRNKEVKAEPKRLVARSLLNQEIRNMQMSFANDVLTUTJFN  
SASCPFIETIENCNLPQRKFP\*

BASE COUNT 278 a 248 c 202 g 346 t

ORIGIN

alignment\_scores:

Quality:	1319.00	Length:	314
Ratio:	4.517	Gaps:	1
Percent Similarity:	92.994	Percent Identity:	82.484

alignment\_block:

US-09-755-017-2 x HSA302589

Align seg 1/1 to: HSA302589 from: 1 to: 1074

1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGlyPhe 17  
|||||  
1 ATGAATGGGTAATGAAGTGTCCACAGAGAGTTCATCTGTAGTTT 50  
17 eSerAspArgProTyrLeuGluPheProLeuValValPheLeuIle 34  
|||||  
51 CTGAGATCAACCATGGCTAGAGATTCACCCCTTGTGATGTTCTGTTT 100  
34 eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50  
|||||  
101 CCTATCTTGTGACATCTTTGGCAATGTGCAATTAATCTTGTCATCAT 150  
51 LeuAspThrLysLeuHisThrPrometLysPhePheLeuThrAsnLeu 67  
|||||  
151 GTGAGTTTCAAACTCCACACCCCTATGACTTTTCTTATGCAATCTCTC 200  
67 rLeuLeuAspLeuGlyTyrThrLysThrValProGlnMetLeuValA 84  
|||||  
201 ACCTCGGAGCTTGTGCTATACACAGTACAGTTCCACAATCTCGTAA 250  
84 sLeuGlySerIleArgLysValIleSerTyrArgLysValAlaGln 100  
|||||  
251 ACATATGCAACACACAGAGATATAGTTATGTGCTGTGCTGCTGCTG 300  
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117  
|||||  
301 CTTTCAATTTCTCGGCTTGGCTTCCACAGATGTCTCTCTGCGCGCT 350  
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 134  
|||||  
351 CATGCTCTTGTATAGCTTGTACCTATTGTCGCGCTCTCCATTTACTCA 400  
134 alIleMetHisGlnArgLysLeuGlnLeuAlaIleAlaSerTyrVal 150  
|||||  
401 TTATCATGCAACACAGAGCTCTGCTTCAGTGTGACACTGATCTGATTT 450  
151 ThrGlyPheSerAsnSerValTyrLeuSerThrLeuThrLeuGlnLeuP 167  
|||||  
451 AGAGCTTTAGCAATTCAGATTAACAGTCCACCTGACACTTAAGATGCC 500  
167 oLeuGlyAspProTyrValIleAspHisPheLeuGlyValProAla 184  
|||||  
501 ACTGTGTGTCAACAAAGAGTGATCATCTTCTGTGAAGTCCCTGCTC 550  
184 eLeuLeuLysLeuSerCysValGluThrThrAlaAsnGlnAlaGluLeuP 200  
|||||  
551 TGTCTCAAGTTGCTGTGTGACACACAGCAATAGGCTGACATTAATTC 600  
601 TTCATCAGTGTGCTATTCTCTTAATACCGCTGACACTCATCTTATATC 650  
217 rTyrAlaPheIleValAlaValAlaLeuArgIleGlnSerAlaGluGlyA 234  
|||||  
651 GATATGCTTTATTTGTCCAAAGAGTGTGAGAAATCCAGTCTGCGAAGTC 700

seq\_name: gp-pr:HSA302590

seq\_documentation\_block:

LOCUS HSA302590 1074 bp DNA PRI 06-JAN-2001  
DEFINITION Homo sapiens 6M1-10\*01 gene for olfactory receptor, cell line YAR.  
ACCESSION AJ302590  
VERSION AJ302590.1 GI:12054404  
KEYWORDS 6M1-10\*01 gene; olfactory receptor.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
uchanska-Ziegler, B., Volz, A., Younger, R. and Beck, S.  
Polymorphic olfactory receptor genes and HLA loci constitute  
extended haplotypes  
(in) Kasahara, M. (Ed.):  
MAJOR HISTOCOMPATIBILITY COMPLEX- EVOLUTION, STRUCTURE, AND  
FUNCTION: 110-130;  
Springer-Verlag, Tokyo, Japan (2000)

REFERENCE  
AUTHORS Ziegler, A., Ehlers, A., Forbes, S. A., Trowsdale, J., Volz, A.,  
Younger, R. M. and Beck, S.  
Polymorphisms in olfactory receptor genes: a cautionary note  
Unpublished  
3. (bases 1 to 1074)  
Younger, R. M., Amadou, C., Bethel, G., Ehlers, A., Fischer, Lindahl, K.,  
Forbes, S. A., Horton, R., Milne, S., Mungall, A. J., Trowsdale, J.,  
Volz, A., Ziegler, A. and Beck, S.  
Characterisation of clustered MHC-linked Olfactory Receptor Genes  
in Human and Mouse  
Unpublished  
4. (bases 1 to 1074)  
Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A., Younger, R.

REFERENCE  
AUTHORS Ehlers, A., Beck, S., Forbes, S. A., Trowsdale, J., Volz, A.,  
Younger, R.  
MHC-linked olfactory receptor loci exhibit polymorphism and  
contribute to extended HLA/OR-haplotypes  
Unpublished  
5. (bases 1 to 1074)  
Ehlers, A.  
Direct Submission  
Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY

JOURNAL  
TITLE

JOURNAL  
AUTHORS

JOURNAL  
AUTHORS

JOURNAL  
AUTHORS

JOURNAL  
AUTHORS

FEATURES

source

1. 1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/cell\_line="YAR"  
1. 1074  
gene

1	MeAsnTrpValAsnAspSerIleIleGlnIuphelleuLeuLeuGlyP	17
1	ATGAATTTGGGTAAATAGAGTGTCCACAGAGGATTCAATCGTAGTTT	50
17	eSerAspArgProTrrPleuGluPheProleuLeuValAlaPheLeuIles	34
51	CTCAGATTCACACCTGGCTGTGAGATTCTCCACCCCTTGTGTGATTTCTGTTT	100
34	eryThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg	50
101	CCATATATTGGACAAATCTTGGCAATCTGCACAAATATTCTTGTTCCAT	150
51	LeuAspThrLysLeuHisThrPrometLyrPhePheLeuThrAsnLeuSe	67
151	GTGAGATTCAAACTCCACACCCCTTATGTACTTTTCTTACAAATCTCTC	200
67	rIleuLeuAspLeuCysTyrThrThrcCysThrValProGlnIleuValAla	84
201	ACTCTCGGACCTTGGTATACCAACAAGTACGTTCACAAATVGGCTGGTAA	250
84	sIeuCysSerIleatrgLysValIleSerTyrArgGlyCysValAlaGln	100
251	ACATATGGCAACACACAGAAAGTAATCAGTTATGGTGGCGCTGTGTGGCCAG	300
101	LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrIleuLeuLeuAlaVala	117
301	CTTTTCATTTTTCTGGCCCTTGGGGTTCACAGAAATGTCTTCTCCGGGCCGT	350
117	IleMetSerPheAspArgPheValAlaIleCysArgProIleuHisTyrSerV	134
351	CATGTGCTTGGATAGAGTTTGTAGCTATTTTGTGGCCCTCCATTACTCAA	400
134	AlleMetHisGlnArgLeuCysLeuGlnLeuAlaAlaAlaSerTrpVal	150
401	TTTATCATGCCACCCAGAGGCTCTGCTTCAGTTGGGACGTGCATCCTGGATT	450
151	ThrGlyPheSerAsnSerValIrrPheuSerThrLeuThrIleuGlnLeuP	167
451	AGTGGCTTTTAGCAATTCAGATTATACAGTCCACCTGGACACTTAAGAGTCC	500
167	oleCysAspProTyrValIleAspHisPheLeuGlyGluValProAlat	184
501	ACTGTGGTGCACAAAGAAGTGGATCAGTCTGTCTGTGAATCCCTGCTC	550
184	euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe	200

## AUTHORS

## JOURNAL

Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik,  
Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin,  
GERMANY

## FEATURES

Location/Qualifiers

1..1074

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="6"

/cell\_line="OLGA"

1..1074

/gene="6M1-10\*01"

1..1074

/gene="6M1-10\*01"

1..1074

/product="Olfactory receptor"

/protein\_id="CAC20511.1"

/db\_xref="GI:12054407"

/translation="MNWYKSVPOEFILLVPSDPELPEPFVMEFLSYILITFENLT  
ILVSHVDFKLHTPMYFELSNLSLDLYTSVPMQMLNICKRVISYSGVAF  
IFLAGSTECILLAWMCDREVAICRPHYSIIMHRCFOLAASWISGFSNVL  
TWTLMPLCGHKEVDHFECEVPALIKISCDPTNEAEFLISVILLIPVLLISY  
APIQAVIRIOSAEORAKRSTGSHLIVSLFSTGTAISMILQPPSPSKRGMVSL  
FCGIIAPLNLPLITLIRNKEVKEAKRLVANSLLNQETRMOMISFADVLITLTFN  
SASCPFIETIENCNLPORKEP"

BASE COUNT 278 a 248 c 202 g 346 t

ORIGIN

## alignment\_scores:

Quality: 1319.00 Length: 314  
Ratio: 4.517 Gaps: 1  
Percent similarity: 92.994 Percent identity: 82.484

## alignment\_block:

US-09-755-017-2 x HSA302591 ..

Align seg 1/1 to: HSA302591 from: 1 to: 1074

```

1 MetAsnTrpValAsnAspSerIleIleGluPheIleLeuGlyPh 17
|||||
1 ATGAATTTGGTAAATAAGAGTGTCCACAGAGAGTCTCTGTAGTTT 50
17 eSerAspArgProTrpLeuGluPheProLeuLeuValPheLeuIle 34
|||||
51 CTCAGATTCACACAGCTAGATTCACCCCTTGATGATCTCTGTTT 100
51 eSerAspArgProTrpLeuGluPheProLeuLeuValPheLeuIle 34
|||||
34 eTrpThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
|||||
101 CCTATATCTTGGACAAATCTTGGCAATCTGACAAATAATCTTGTGCACAT 150
101 LeuAspThrLysLeuHisThrPrometLysPhePheLeuThrAsnLeu 67
|||||
151 GTGGATTTCAACTCCACACCCCTATGTACTTTTCTTACCAATCTCTC 200
151 rLeuLeuAspLeuGlyThrThrCysThrValProGlnMetLeuValA 84
|||||
201 ACTCCGAGACCTTGTCTATACACAGTACAGTTCACAAATCTCTGTAA 250
201 snLeuGlySerIleArgLysValIleSerTyrArgGlyCysValAlaGln 100
|||||
251 ACATATGCAACACGAGAAAGTAAATCATGTATGTGTGTGTGGCCAG 300
251 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAla 117
|||||
301 CTTTATATTTCTCGGCTTGGCTTCCACAGAAATGCTTCTCTGCGCGT 350
301 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerY 134
|||||
351 CATGTCTTGTAGATGTTGTACTATTTGTGCGCTCTCCATTAATCAAA 400
351 aIleMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaSerTrpVal 150
|||||
401 TTATCATGACACGAGAGCTCTGCTTCCAGTTGACAGCTGATCTGATT 450

```

```

151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuPr 167
|||||
451 AGTGGCTTTAGCAATTCAGTATTCAGATCCACCTGGACACTTAAGATGCC 500
167 OleuGlyAspArgProTyrValIleAspHisPheLeuGlyValAlaProAla 184
|||||
501 ACTGTGTGTGCACAAAGAGTGCATCTTCTGTGAGTGCCTGCTC 550
184 euleuGlyLeuSerGlyValAlaGluThrThrAlaAsnGluAlaGluLeuPhe 200
|||||
551 TGCTCAAGTGTCTCTGTGTGACACACACCAATGAGGCTGACATATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuLeuLeuIle 217
|||||
601 TTATATCGTGTCTATTCCTTAAATACCGGTACACTCATCTTATATC 650
217 rTyraIaPheIleValAlaGluValLeuArgIleGlnSerAlaGluGlyA 234
|||||
651 GTATGCTTTATTTGTCACACAGCTGTTGACAAATCCAGTCTGGAAGTTC 700
234 rGluIleValAlaPheGlyThrCysGlySerHisLeuIleValAlaSerLeu 250
|||||
701 AACGAAAGGCAATTTGGGACATGTGGCTCCCATCTAATTGGTGTACATT 750
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267
|||||
751 TTTTATGGTACAGCTATCTCAGTACCTGACCAACCTTCACCAAGCTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284
|||||
801 CAAAGACCGGGGAAAGATGTTCTCTCTCTGGAATCATTCACACCCA 850
284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
|||||
851 TGCTGATCCCTTATATATACCTTAAAGCAAAAGGAAAGGAAAGCC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
|||||
901 TTTTAAAGGTTGGTTCCAAGAGAGTCTTCTTAATCAAGAAA 940
seq_name: gb_pr:HSA302592
seq_documentation_block: 1074 bp DNA PRI 06-JAN-2001
LOCUS HSA302592
DEFINITION Homo sapiens 6M1-10*01 gene for olfactory receptor, cell line AMAT.
ACCESSION AJ302592
VERSION AJ302592.1 GI:12054408
KEYWORDS 6M1-10*01 gene; olfactory receptor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Uchanska-Ziegler,B., Volz,A., Younger,R. and Beck,S.
Polymorphic olfactory receptor genes and HLA loci constitute
extended haplotypes
(in) Kasahara,M. (Ed.);
MAJOR HISTOCOMPATIBILITY COMPLEX-EVOLUTION, STRUCTURE, AND
FUNCTION: 110-130;
Springer-Verlag, Tokyo, Japan (2000)
REFERENCE
Ziegler,A., Ehlers,A., Forbes,S.A., Trowsdale,J., Volz,A.,
Younger,R.M. and Beck,S.
Polymorphisms in olfactory receptor genes: a cautionary note
unpublished
3 (bases 1 to 1074)
JOURNAL
Younger,R.M., Anadou,C., Bethel,G., Ehlers,A., Fischer Lindahl,K.,
Forbes,S.A., Horton,R., Milne,S., Mungall,A.J., Trowsdale,J.,
Volz,A., Ziegler,A. and Beck,S.
Characterisation of clustered MHC-linked Olfactory Receptor Genes
in Human and Mouse
TITLE

```



JOURNAL Unpublished  
 REFERENCE 4 (bases 1 to 1074)  
 AUTHORS Ehlers, A., Beck, S., Forbes, S.A., Trowsdale, J., Volz, A., Younger, R. and Ziegler, A.  
 TITLE MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA/OR-haplotypes  
 JOURNAL Unpublished  
 REFERENCE 5 (bases 1 to 1074)  
 AUTHORS Ehlers, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-NOV-2000) Ehlers A., Institut fuer Immunogenetik, Universitaetsklinikum Charite, Spandauer Damm 130, 14050 Berlin, GERMANY

# FEATURES

Source Location/Qualifiers  
 1..1074  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /cell\_line="AMA1"  
 1..1074  
 /gene="6M1-10\*01"  
 1..1074  
 /gene="6M1-10\*01"  
 /product="Olfactory receptor"  
 /codon\_start=1  
 /protein\_id="CAC20512.1"  
 /db\_xref="GI:12054409"  
 /translation="MNWVKSVPQEFILVFSDDPWEIPPEWFLSYILITFQNLTLVSHVDFKLHTPMYFPLNLSLDLCYTTSTYQPMVINCNRKVISYGCVAOLFIFLAGSTECFLAVMCFDRFAICRPLHYSIIMHRCFOLAASWISGFSNSVLOS TWTLMPLCGHREVDHFECEVDALIKSCVDTPANEALFISVFLIPYLLIISYAFIVAOVARIOSASGQRFCGSHLIVSLFPGTASMTLOPSPSKRGKWSLFCGIIAPMLNPLITLIRNKEAFKRLVAASLNLQETRMQMSIFADTVLTITLNF SASCPITFITTENCNLPQRKFP"

BASE COUNT 278 a 248 c 202 g 346 t  
 ORIGIN

alignment\_scores:  
 Quality: 1319.00 Length: 314  
 Ratio: 4.517 Gaps: 1  
 Percent Similarity: 92.994 Percent Identity: 82.484

alignment\_block:  
 US-09-755-017-2 x HSA302592 ..  
 Align seg 1/1 to: HSA302592 from: 1 to: 1074

```

1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleuLeuGlyPh 17
|||||
1 ATGAATTGGTAAATTAAGAGTGTCCACAGAGAGTTCAATTCGTAGTTT 50
17 eSerAspArgProTrpLeuGluPheProLeuLeuValPheLeuIles 34
|||||
51 CTCAGATCAACACATGGCTAGAGATCCACCTTGTGATCTTCGTGTTT 100
51 erTyThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
|||||
101 CCTATATCTTGACAAATCTTGGCAATCTGCAATAATCTTGTGTCCAT 150
51 LeuAspThrIleLeuHisThrPrometIlePhePheLeuHisAsnLeu 67
|||||
151 GTGGATTTCACAACTCCACACCCCTATGTACTTTTCTTGAACAATCTTC 200
67 rIleuLeuAspLeuGlySerIleThrIleGlyThrValProGlnMetLeuVal 84
|||||
201 ACTCCGAGACCTTGGCTATACACAGATGATCCCAATGCTGTGTTA 250
84 snLeuGlySerIleArgIleValIleSerIleArgIleGlyValAlaGln 100
|||||
251 ACATATGCAACACACGAAAGTATGATGCTGTGTGTGCTGTGTGCTG 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluThrIleLeuLeuAlaVal 117

```

```

|||||
301 CTTTCAATTTCTCGGCTGGGTTCCACAGAAATGTCTTCTCGGCGCT 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisIleTySerV 134
|||||
351 CATGTGCTTGAATAGATTGTAGCTATTTGTGGGCTCTCCATATACCA 400
134 alIleMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaIleSerTrpVal 150
|||||
401 TTATCATGACACAGAGGCTGCTGCTCAGTTGGCAGCTGCATCCGATT 450
151 ThrGlyPheSerAsnSerValTrpLeuSerThrIleuGlnLeuPhe 167
|||||
451 AGTGGCTTTAGCAATTCAGATTACAGTCCACCTGGACACTTAAGATGCC 500
167 oLeuGlyAspProTrpValIleAspHisPheLeuGlyValProAlaL 184
|||||
501 ACTGTGTGTCACAAAGAGTGCATCTTCTGTGAGAGTCCGCTGCTC 550
184 euleuGlySerIleuGlyValGluThrThrAlaAsnGluAlaGluPhe 200
|||||
551 TGCTCAAGTTCCTGCTGTGACACACAGCAATGAGGCTGAACTATTC 600
201 leuValSerGluLeuPheHisIleuIleProLeuThrIleuLeuIleSe 217
|||||
601 TTCAATCATGCTGCTATTCCTTCTTAATACCCGTACACTCATCTTATATC 650
217 rTyAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyA 234
|||||
651 GTATGCTTTTATTTGTCACACAGTGTGAAATCCAGTGTGCGAAGGTC 700
234 rGlnGlnAlaIlePheGlyThrCysGlySerHisIleuIleValSerLeu 250
|||||
701 AACGAAAGGATTTGGACATGTGGCTCCCATTAATTTGGGTGCTACT 750
251 PheTySerThrAlaValSerValTyLeuGlnProProSerProSerSe 267
|||||
751 TTTTATGATGATACAGTATCTCATGATGATCTGCAACACCTTCACCCAGCTC 800
267 rIlyAspGlnGlyLysMetValSerLeuPheTyGlyIleIleAlaPhe 284
|||||
801 CAAGAAGCCGGGAAAGATGTTCTCTCTCTGTAATCATTTGACACCA 850
284 eIleuAsnProLeuIleTyThrIleuArgAsnLysGluValLysGlnGly 300
|||||
851 TGCTGAATCCCTTATATATACACTTGGAACAAAGGTTAAAGGAAGCC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
|||||
901 TTTRAAAGGTTGCTTGCAAGAGCTCTTATATCAAGAAA 940
seq_name: gb_pr:HS193B12
seq_documentation_block:
LOCUS HS193B12 100375 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from PAC 193B12 on chromosome 6p21.3-22.3.
Contains histone H2A/d, H2B/d, H2A.1, H1.5, H2.F, H4.k, H3.J
genes, histone h2a.1 and h2b.2 duplicate genes, glycine (GCG) tRNA gene,
olfactory receptor OL1 like gene, ESTs SRS and predicted CpG
islands.
ACCESSION Z98744
VERSION Z98744.1 GI:3080457
KEYWORDS 6p21.3-22.3; A4; CpG island; glycine tRNA; H1.5; H2A; H2A.1;
H2A/d; H2B; H2B.2; H2B/d; H3; H3.1; H3.F; H3/j; H4; H4.k; histone;
OL1; olfactory receptor; ph2b.1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 100375)
AUTHORS Wild, A.
TITLE Direct Submission

```

## JOURNAL

## COMMENT

Submitted (12-DEC-1997) Chromosome 6 Project Group  
 (http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Apr 24, 1998 this sequence version replaced g1:2578067.  
 IMPORTANT: This sequence is the entire insert of clone 193B12.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variations annotated may not be found in the sequence submission  
 corresponding to the overlapping clone as we submit sequences with  
 only a small overlap as described above.  
 This sequence was generated from part of bacterial clone contigs of  
 human chromosome 6, constructed in collaboration by the Sanger  
 Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler,  
 David Ruddy, Jeffrey Gruen.

Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6/ This sequence has been finished  
 according to sequence map criteria as follows. An attempt is made  
 to resolve all sequencing problems, such as compressions and  
 repeats, but not necessarily within known annotated human repeat  
 sequence elements (e.g. Alu). Where the sequence is ambiguous,  
 there is an annotation using the 'unsure' feature key.  
 The true left end of clone 193B12 is at 1 in this sequence. The  
 true right end of clone 193B12 is at 100375.  
 193B12 is from the library Rpci1 constructed at the Roswell Park  
 Cancer Institute by the group of Pieter de Jong.  
 For further details see http://bacpac.med.buffalo.edu/.

## FEATURES

## source

1..100375  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6p21.3-22.3"  
 /clone="Rpi-193B12"  
 /clone\_11b="Rpci-1"  
 211..312  
 /note="MIR2 repeat: matches 33. .137 of consensus"  
 534..768  
 /note="Alu repeat: matches 289. .35 of consensus;  
 incomplete repeat"

repeat\_region  
repeat\_region

## prim\_transcript

## mRNA

/complement(1001..1308)  
 /note="match: 5' EST H54606"  
 complement(<1589..>2087)  
 /gene="H2A/d"

gene  
CDSmisc\_feature  
mRNA

/note="match: ESTs AA426351 AA188780 AA145892 AA048695  
 AA061654 H15508 W41535 AA272868 AA453105 W65215 AA231513  
 AA073767 AA528953 AA492640 W11262 AA060372 AA027719  
 AA143419 G90551; match: cDNAs L19778 L19779 X58069 X14850  
 D17284; match: genomic DNAs Z83739 X83549 Z83736 Z83742  
 X05862 Y00117 X16148 Z30940 U62674 X57138 U62669 X02218  
 U62673 X07763 X07758"  
 /evidence="not\_experimental"  
 /product="histone H2A"  
 complement(1589..2087)  
 /gene="H2A/d"  
 complement(1688..2080)  
 /gene="H2A/d"  
 /codon\_start=1  
 /evidence="not\_experimental"  
 /product="histone H2A"  
 /protein\_id="CAB11417.1"  
 /db\_xref="GI:3080458"  
 /translation="MSGRGKGGKARAKAKTRSSRAGLPVGRVHRLRKGNVAERY  
 GAGAPVLAIVLEVTATLEAGNAARDNKKRTIIPHLQAIANDLEMLKLGKVT  
 IAQGVLPNTQIVLPLPKTESHHKKGK"  
 1807..2816  
 /note="putative CPG island"  
 <2327..>2857  
 /gene="H2B/d"  
 /note="H2B.1 and H2B.2 like; match: ESTs AA354038 AA217829  
 W49399 AA396588 AA080125 W41523 AA218213 AA423665 AA562098  
 AA493023 AA237851 W41584 AA255145 AA327803 AA064568

gene  
CDS

2327..2857  
 /gene="H2B/d"  
 2403..2783  
 /gene="H2B/d"  
 /note="H2B.1' and H2B.2 like; match: proteins Q00715 P10854  
 P33778 P02278 P02294 Q64477 CE07075 P27795 CE05165 CE07280  
 P10853"  
 /codon\_start=1  
 /evidence="not\_experimental"  
 /product="histone H2B"  
 /protein\_id="CAB11418.1"  
 /db\_xref="GI:3080459"  
 /db\_xref="SPTREMBL:Q99877"  
 /translation="MPESKSPAPKGGSKAVYKAOKKDGKRRKRRESYVYVK  
 VLKQVPTGTGTSKAMGKIMNSFVNDIFRIAGEASRLAHYKRRSTITSREIQTAVRL  
 LPEGLAKHAYSEGGKAVYKRTSSK"  
 2956..3243  
 /note="Alu repeat: matches 13. .303 of consensus"  
 4365..4499  
 /note="Alu repeat: matches 1. .134 of consensus;  
 incomplete repeat"  
 4501..4556  
 /note="Alu repeat: matches 1. .134 of consensus;  
 incomplete repeat"  
 4786..5085  
 /note="Alu repeat: matches 2. .300 of consensus"  
 6646..6723  
 /note="Alu repeat: matches 1. .1 of consensus"  
 7038..7077  
 /note="39 copies of 2 mer 82 & conserved"  
 /note="20 copies of 2 mer 85 & conserved"  
 7408..7702  
 /note="Alu repeat: matches 3. .297 of consensus"  
 10347..10644  
 /note="Alu repeat: matches 2. .300 of consensus"  
 11933..11977  
 /note="MIR2 repeat: matches 100. .146 of consensus"  
 12051..12094  
 /note="22 copies of 2 mer 98 & conserved"  
 12459..12640  
 /note="Alu repeat: matches 120. .299 of consensus;  
 incomplete repeat"  
 12834..13127  
 /note="Alu repeat: matches 1. .296 of consensus"  
 15893..16409  
 /note="Alu repeat: matches 1. .296 of consensus"  
 17170..17463  
 /note="match: BAC end B47865"  
 17841..18149  
 /note="Alu repeat: matches 293. .1 of consensus"  
 18070..18215  
 /note="match: 5' ESTs D31045 D31402"  
 18695..18995  
 /note="2 copies of 73 mer 95 & conserved"  
 19116..19178  
 /note="Alu repeat: matches 1. .302 of consensus"  
 20099..20194  
 /note="MIR repeat: matches 173. .110 of consensus"  
 20197..20490  
 /note="MIR repeat: matches 1. .95 of consensus"  
 20495..20572  
 /note="Alu repeat: matches 301. .8 of consensus"  
 20574..20881  
 /note="MIR repeat: matches 84. .161 of consensus"  
 20912..20996  
 /note="Alu repeat: matches 308. .1 of consensus"  
 /note="FLAM\_C repeat: matches 1. .85 of consensus"

```

repeat_region 21030..21345
/Note="MT1C repeat: matches 169..466 of consensus"
repeat_region 22763..23053
/Note="MER2 repeat: matches 345..19 of consensus"
repeat_region 23261..23657
/Note="MT1F repeat: matches 480..44 of consensus"
repeat_region 23759..24062
/Note="AluSp repeat: matches 1..303 of consensus"
repeat_region 25062..25189
/Note="MIR2 repeat: matches 2..137 of consensus"
repeat_region 26058..26355
/Note="AluX repeat: matches 302..8 of consensus"
repeat_region 27721..27772
/Note="2 copies of 26 mer 96 % conserved"
gene complement(27803..28142)
/Note="2 copies of 26 mer 96 % conserved"
CDS /gene="pH2B.1"
complement(27803..28142)
/Note="pH2B.1"
/Note="match: proteins Q90859 P06899 P10854 P23527 P33778
O64524 P06900 P02279 P02278 P02281 P10853 O64477: match:
ESRs AA256738 AA327803 AA286933 AA315768 AA286932 AA280257
AA493453 AA423665: match: CDNAS M18046 M18045 M69756
M23487: match: genomic DNAs X83547 Z83336 Z80780 U62669
X59961 U77510 X57138"
/codon_start=1
/pseudo
/evidence=not_experimental
misc_feature 27829..29369
/Note="putative Cpg island"
repeat_region 28190..28492
/Note="AluX repeat: matches 1..302 of consensus"
repeat_region 28508..28807
/Note="AluSg repeat: matches 2..303 of consensus"
mRNA <29070..>29531
/Note="H2A.1"
/Note="H2A.1"

alignment_scores:
Quality: 1319.00 Length: 314
Ratio: 4.517 Gaps: 1
Percent Similarity: 92.994 Percent Identity: 82.484

alignment_block:
US-09-755-017-2 x HS193B12/rev
Align seg 1/1 to reverse of: HS193B12 from: 1 to: 100375

1 MetAsnTPPVAlAsnAspSerIleIleGlnGluPheIleLeuGly 17
76061 ATGAATGGGTAAATAGAGAGTCCACAGGAGTTCCTGTACTTTT 76012
17 eSerAspArgProTyrPheGluPheProLeuValValPheLeu 34
76011 CTCAGATTCACACATGGCTAGAGATTCACCTTTGGATGTTCTGTTT 75962
34 eTyrThrValThrIlePheGlyAsnLeuThrIleLeuValSerArg 50
75961 CCTATATCTTGCAATCTTTGGCAATCTGACATTAATCTTGTCACAT 75912
51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeu 67
75911 GGGAAATTCAAACCTCCACACCCCTATGTAATCTTTCTTAGCAATCTGC 75862
67 rLeuAsnAspLeuGlyTyrThrThrCysThrValProGlnMetLeuVal 84
75861 ACTCCGGACCTTTGCTATACACCAAGTACAGTCCACAAAGCTGTAA 75812
84 snLeuCysSerIleArgLysValIleSerTyrArgLysValAlaGln 100
75811 ACATATGCACACACGAAAGATATAGTTATGCTGCTGTGCCCCAG 75762
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117
75761 CTTTCATTTTCTGCGCTTGGGTTCCACAGATGCTCTCTGCGCGCT 75712

```

```

117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 134
75711 CATGTCCTTTGATAGAGTTGTACTATTGTGCGCTCTCCATTAATCA 75662
134 aIleMetHisGlnArgLysCysLeuGlnLeuAlaAlaSerTyrVal 150
75661 TTATCATGCACAGAGAGCTGCTGCTCCAGTTGGCAGCTGCATCGAT 75612
151 ThrGlyPheSerAsnSerValIrrPheLeuSerThrIleLeuGlnLeu 167
75611 AGGCGCTTAGCAATTCAGATTACAGTCACCTGACACTTAAGATGCC 75562
167 oLeuCysAspProTyrValIleAspHisPheLeuGlyValProAla 184
75561 ACTGTGTGCTCCAAAGAGTGATCTTCTGTGAGTCCGTGCTC 75512
184 euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeu 200
75511 TGCTCAAGTTGCTCTGTGACACAAACAGCAATGAGCTGACTATTC 75462
201 leuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeu 217
75461 TTCATCAGTGTGCTATTCTTAAATACCCGTGACACTCATCTATATC 75412
217 rTyrAlaPheIleValAlaValAlaLeuArgIleGlnSerAlaGluGly 234
75411 GTATGCTTTTATTTGTCACAGCACTGTGTGAGAACTGCTGAGAGGTC 75362
234 rGlnLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 250
75361 AAGCAAGCATTTGGGACATGTCCTCCATTAATGTGGTGTACTT 75312
251 PheTyrSerThrAlaValSerValTyrLeuGlnProSerProSer 267
75311 TTTTATGATGACAGCTATCTCCATGCTACGCAACACCTTACCCAGCTC 75262
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284
75261 CAAGACCGGGGAAAGATGTTCTCTCTGTAATCATTCACCA 75212
284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGlu 300
75211 TGCTGATCCCTTATATATACCTTAGGACAAAGTAAGGAAGCC 75162
301 PheLysArgLeuValAlaArgValPheLeuIleLys 313
75161 TTAAAGTTGGTGCACAAAGCTCTTAAATCAAGAA 75122

seq_name: gb_hlg:AC024428

seq_documentation_block:
LOCUS AC024428 166758 bp DNA HTG 26-MAY-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-59801 map 3, WORKING DRAFT
SEQUENCE 17 unordered pieces.
ACCESSION AC024428.3 GI:8076917
VERSION AC024428.3
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 166758)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-59801
JOURNAL Unpublished
2 (bases 1 to 166758)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRellano,K., Dewar,K., Diaz,J.S., Dodde,S., Domingo,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,

```

TITLE  
JOURNAL  
COMMENT

Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Largoczek, K., Lehoczyk, J., Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marguis, N., McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Mhova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirelli, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Ye, W. J., Young, G., Zahoun, J., Zimmer, A. and Zody, M.

## Direct Submission

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced g1:727173.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

## Project Information

Center project name: L6934

Center clone name: 598\_O\_1

## Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 16318 bases at least Q40

Consensus quality: 161224 bases at least Q20

Consensus quality: 163451 bases at least Q20

Insert size: 191000; agarose-fp

Insert size: 165158; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1 1025: contig of 1025 bp in length
* 1026 1125: gap of 100 bp
* 1126 3025: contig of 1900 bp in length
* 3026 3125: gap of 100 bp
* 3126 4346: contig of 1221 bp in length
* 4347 4446: gap of 100 bp
* 4447 7420: contig of 2974 bp in length
* 7421 7520: gap of 100 bp
* 7521 10669: contig of 3149 bp in length
* 10670 10769: gap of 100 bp
* 10770 14258: contig of 3489 bp in length
* 14259 14358: gap of 100 bp
* 14359 18958: contig of 4600 bp in length
* 18959 19058: gap of 100 bp
* 19059 23968: contig of 4910 bp in length
* 23969 24068: gap of 100 bp
* 24069 28546: contig of 4478 bp in length
* 28547 28646: gap of 100 bp
* 28647 37679: contig of 9033 bp in length
* 37680 37779: gap of 100 bp
* 37780 45607: contig of 7828 bp in length
* 45608 45707: gap of 100 bp
* 45708 54439: contig of 8732 bp in length

```

```

* 54440 54539: gap of 100 bp
* 54540 63730: contig of 9191 bp in length
* 63731 63830: gap of 100 bp
* 63831 77842: contig of 14012 bp in length
* 77843 77942: gap of 100 bp
* 77943 98659: contig of 20717 bp in length
* 98660 98759: gap of 100 bp
* 98760 127843: contig of 29084 bp in length
* 127844 127943: gap of 100 bp
* 127944 166758: contig of 38815 bp in length.
Location/Qualifiers
1. 166758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-59801"
/clone_lib="RP11 Human Male BAC"
1. 1025
/note="assembly_fragment"
1126. 3025
/note="assembly_fragment"
3126. 4346
/note="assembly_fragment"
clone_end:77
vector_side:left"
4447. 7420
/note="assembly_fragment"
7521. 10669
/note="assembly_fragment"
10770. 14258
/note="assembly_fragment"
14359. 18958
/note="assembly_fragment"
19059. 23968
/note="assembly_fragment"
24069. 28546
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
28647. 37679
/note="assembly_fragment"
37780. 45607
/note="assembly_fragment"
45708. 54439
/note="assembly_fragment"
54540. 63730
/note="assembly_fragment"
63831. 77842
/note="assembly_fragment"
77943. 98659
/note="assembly_fragment"
98760. 127843
/note="assembly_fragment"
127944. 166758
/note="assembly_fragment"

```

```

BASE COUNT 48778 a 33823 c 33960 g 48591 t 1606 others
ORIGIN

```

```

alignment_scores:
Quality: 1319.00 Length: 314
Ratio: 4.517 Gaps: 1
Percent Similarity: 92.994 Percent Identity: 82.484

```

```

alignment_block:
US-09-755-017-2 x AC024428

```

```

Align seg 1/1 to: AC024428 from: 1 to: 166758

```

```

1 MetasnrTpyValasnspsrlelleGIngluphelleuLeuGlyph 17
|||||
99340 ATGAATTGGGTAAATAAAGAGTGTCCACAGAGATTCATTCTGTACTTTT 99389

```

[illegible]

```

seq_name: gb|btg:AC025941

seq_documentation block:
LOCUS      AC025941      176277 bp      DNA      26-MAY-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-635011 map 6, WORKING DRAFT
SEQUENCE   AC025941      15 unordered pieces.
ACCESSION   AC025941
VERSION     AC025941.2      GI:7408057
KEYWORDS    HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 176277)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 6, clone RP11-635011
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 176277)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```

TITLE	Direct Submission
JOURNAL	Submitted (17-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Apr 4, 2000 this sequence version replaced gi:7259782.

----- Genome Center  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

```
----- Project Information -----
sequence_submissions
```

Project Information  
Center project name: L8577

Center clone name: 635\_0\_11

## ----- Summary Statistics

Sequencing vector: M13; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% conc  
Assembly program: Phred; version 0.960721

Consensus quality: 168929 bases at least 94%  
assembly program: Phrap, version 0.300731

Consensus quality: 172827 bases at least 0300

Consensus quality: 174151 bases at least Q20

Insert size: 176000; agarose-fp

Insert size: 174877; sum-of-contrigs

Quality coverage: 4.5 In Q20 bases; agarose-1p  
Quality coverage: 4.6 In Q20 bases; sum-of-confrs

guaranty coverage. 7.0 in 220 bases/ 5000 of coverage

NOTE: This is a 'working draft' sequence. It currently

consists of 15 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

his record will be updated with the finished sequence

soon as it is available and the accession number will be updated with the finished sequence.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-635O11"
/clone_id="RPCI-11 Human
1..2997
/note="assembly_fragment"
3098..5606
/note="assembly_fragment
clone_end:SP6
vector_side:left"
5707..9209
/note="assembly_fragment"
9310..13614
/note="assembly_fragment"
13715..20639
/note="assembly_fragment"
20740..26545
/note="assembly_fragment
clone_end:T7
vector_side:right"
26646..37068
/note="assembly_fragment"
37169..48083
/note="assembly_fragment"
48184..59015
/note="assembly_fragment"
59116..70420
/note="assembly_fragment"
70521..80466
/note="assembly_fragment"
80567..92325
/note="assembly_fragment"
92426..110572
/note="assembly_fragment"
110673..131898
/note="assembly_fragment"
131999..176277
/note="assembly_fragment"

```

1 MetAsnTrpValAsnAspSerIleIleGlnIuphIleLeuLeuGlyPh 17  
41072 ATGAATTTGGTAATAATAGAGTGTCCACAGAGACTTATCTGTAGTTT 41122

17 eSerAspArgProTrpLeuGluPherProLeuLeuValValPheLeuIleS 34  
41122 CTCACATGACACCATGGCTAGAGAAATTCCACCCCTTGTGATGTTCTGTGTTT 41172

34 eTrpThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50  
41172 CCTATATCTGGCAATCTTTGGCAATGTGCACATATATCTTGTGTACAT 41222

51 LeuAspThrIlyLeuIleIsthrProMetGlyrPhePheLeuThrAsnLeuS 67  
41222 GTGGATTTCCAAACGTCCACACCCCTATGACTTTTCTTAGCAATCTCTC 41272

67 rLeuLeuAspLeuGlyCysTyrThrThrCysThrValProGlnMetLeuVal 84  
41272 ACTCTGGACCTTTGCTGTATACACAGAATGACATGTCACAAATGCTGGTAA 41322

84 snLeuGlySerTrpLeuGlyValIleSerTyrArgGlyCysValAlaGln 100  
41322 ACATATGGCAACCCGGAAGAATGACGTATGAGTGGCTGTGTGGCCCAAG 41372

101 LeuPheIlePheLeuAlaLeuGlnGlyAlaThrGlyTyrLeuLeuAlaVal 117  
41372 CTTTTCATTTTCTGTGGCTTGGGTTCCACAGAAATGCTTCTTCTGGCCCT 41422

117 lMetSerPheAspArgPheValAlaIleCysArgProLeuIsthrTyrSerV 134  
41422 CATGGCTTTGATAGTGTGTAGATATTGTTCGGCTCTCCATTCACAA 41472

134 alIleMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaIleAserTrpVal 150  
41472 TTATCATCCACCAAGAGGTCTGTCTCCAGTGGGACGTGCATCTGTGAT 41522

151 ThrGlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuP 167  
41522 AGTGGCTTTAGCAATTCAGTATTACAGTCCACCTGGACACTTAAGATGCC 41572

167 oLeuGlyAspProTyrValIleAspHisPheLeuGlyGluValProAlaL 184  
41572 ACTGTGTGGTCACAAGAGAGTGACACTTCTTGTGTGAAGTCCCTGGTC 41622

184 euLeuLeuLeuSerCysValGluThrThrThrAlaAsnGluAlaGluLeuP 200  
41622 TGCTCAAAATTTGCTGTGTGTGCACAAACAGCAAAATGAGGCTAACTATTTC 41672

201 LeuValSerGluLeuPheHisLeuLeuProLeuThrLeuIleLeuIleS 217  
41672 TTCATCACTGTGCTATTCTCTTATACCCGGAGACACTCATCTTATATC 41722

217 rTyrAlaPheIleValAlaGluAlaValLeuArgIleGlnSerAlaGluGly 234  
41722 GTATGCTTTTATTTGTCCAAGCAGTGTGAGAAATCCAGTCTGCTGAAGTC 41772

234 rGlnGluAlaPheGlyThrCysGlySerHisIleuIleValValSerLeu 250  
41772 AAGCAAAAGCATTTGGGACATGTGGCTCCCACTTAATTTGGTGTACACTT 41822

```
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41822 TTTTATGGTACAGCTATCTCCATGTACCTGCACACCTTCACCCAGCTC 41871
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaProM 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41872 CAAAGACCGGGGAAAGATGGTTTCTCTCTCTGTCGATCATTCACCCA 41921
284 eLLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41922 TGCTGAATCCCTTATATATACACTTAGAACAAAGAGGTAAAGGAGGCC 41971
301 PheLysArgLeuValAla.ArgValPheLeuIleLysLys 313
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
41972 TTTAAAGGTGTGTCAAAGAGCTCTCTTAATCAAGAAA 42011
```





OM of: US-09-755-017-2 to: N\_Geneseq\_1101.\* out\_format: pfs

Date: Feb 26, 2002 5:28 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

```
-MODEL=framed_p2n.model -DEV=xlh
-Q/cgnt_1/USPTO.spool/US09755017/runat.25022002.165641.20808/app.query.fasta_1.374
-DB=N_Geneseq_1101 -QPM=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPEXT=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODEL=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US09755017.@CGN1_1.213 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

#### Search information block:

```
Query: US-09-755-017-2
Query length: 313
Database: N_Geneseq_1101.*
Database sequences: 93621
Search length: 428662619
Search time (sec): 112.140000
```

#### score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation				
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1601.00	2900.23	1.8e-153	939		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1601.00	2900.19	1.8e-153	942		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1601.00	2900.17	3.4e-153	1488		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1319.00	2384.62	9.5e-125	1071		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1319.00	2384.62	9.5e-125	1071		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1138.00	2055.86	1.9e-106	939		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1138.00	2055.86	1.9e-106	939		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				1083.00	1959.59	4.5e-101	648		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				954.00	1720.01	9.9e-88	979		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				941.00	1702.35	9.6e-87	927		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				941.00	1696.81	1.9e-86	933		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				939.00	1693.13	3.1e-86	936		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				918.00	1654.02	4.7e-84	1011		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				912.00	1643.92	1.7e-83	936		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				908.00	1635.88	4.8e-83	1002		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				902.00	1619.82	3.8e-82	1629		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				898.50	1617.70	5.0e-82	1087		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				897.00	1616.54	5.8e-82	936		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				894.00	1610.35	1.3e-81	1002		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				891.00	1605.60	2.3e-81	936		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				889.00	1601.11	4.2e-81	1014		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				889.00	1597.41	6.7e-81	1440		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				883.00	1590.78	1.6e-80	957		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				883.00	1590.78	1.6e-80	957		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				882.00	1590.63	1.6e-80	917		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				879.00	1583.53	4.0e-80	954		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				863.50	1555.49	1.4e-78	933		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				861.00	1550.76	2.7e-78	948		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				861.00	1550.76	2.7e-78	948		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				859.00	1547.12	4.2e-78	948		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				853.00	1536.04	1.8e-77	960		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				849.00	1529.33	4.1e-77	909		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				810.50	1454.20	6.3e-73	1450		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				808.00	1455.78	5.2e-73	810		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				804.00	1446.30	1.7e-72	997		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				801.00	1441.29	3.3e-72	954		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				799.50	1438.62	4.7e-72	948		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				799.50	1438.62	4.7e-72	1000		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				791.00	1437.71	5.3e-72	948		
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018				791.00	1437.71	5.3e-72	921		

/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018 + 790.00 1421.05 4.5e-71 971  
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018 + 787.00 1416.03 8.5e-71 930  
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018 + 787.00 1415.79 8.7e-71 951  
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018 + 786.50 1415.29 9.3e-71 915  
/SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018 + 786.50 1415.26 9.4e-71 918

#### seq\_name: /SIDB2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32018

seq\_documentation\_block:

ID AAH32018 standard; DNA; 939 BP.

AAH32018;

30-JUL-2001 (first entry)

Human olfactory receptor polynucleotide, SEQ ID NO: 591.

Human; olfactory receptor; OR; primary scent determination;

secondary scent determination; polypeptide library; odour receptor;

scent profile; scent fingerprint; scent representation; ds.

Homo sapiens.

WO200127158-A2.

06-OCT-2000; 2000WO-US27582.

08-OCT-1999; 99US-0158615.

24-FEB-2000; 2000US-0184809.

(DIGI-) DIGISCENTS.

(YEDA) YEDA RES & DEV CO LTD.

Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

WPI: 2001-290713/30.

New polynucleotides which encode polypeptides involved in olfactory

sensation for identifying olfactory agonists and antagonists -

Claim 8; Page 417; 1857pp; English.

The present sequence is one of a number of isolated polynucleotides  
which encode polypeptides involved in olfactory sensation. The  
polynucleotides can be used in screening for olfactory agonists and  
antagonists. The methods allow for the determination of primary  
scents and the identification of the odour receptors used to detect  
these primary scents. The methods also enable determination of  
secondary scents and the identification of combinations of odour  
receptors that are involved in detecting such secondary scents.  
This enables the construction of a scent representation (also called  
a scent fingerprint or scent profile), which may be used to re-create  
and edit scents. Libraries of olfactory receptors are useful for  
determining the interaction pattern of a composition with the receptors,  
and can be used for determining differences in the olfactory faculties  
of different individuals.

Sequence 939 BP; 220 A; 227 C; 187 G; 305 T; 0 other;

#### alignment\_scores:

Quality: 1601.00 Length: 313  
Ratio: 5.115 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-755-017-2 x AAH32018 ..  
Align seg 1/1 to: AAH32018 from: 1 to: 939

1 MetasnrPrVaIAsnAspSerIleIleGInGluPhelIeulcLcylph 17

```

|||||
1 ATGAATGGGTAATGACACATCATACAGAGATTATCTCGCGGTTT 50
17 eSerAspArgProTirPLeuGluPheProLeuValValPheLeuIles 34
51 CTGAGATCGACCTTGCTGGAGTTTCCACCTCTTGCTCTCTGATTT 100
34 eTyrThrValAlaThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
101 CTACACTGTGACCATCTTGGCAATGTGACCATTTATCTAGTCTCAGC 150
51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeuSe 67
151 CTGGACACCAACTTCATACCCCATGTATTTTCTTACCAATCTATC 200
67 rLeuLeuAspLeuGlyTyrThrCysThrValProGlnMetLeuValAla 84
201 ACTGCTGGATCTTTGTACACACATGTACAGTCCCAATAGCTACTAA 250
84 snLeuGlySerIleArgLysValIleSerTyrArgGlyCysValAlaGln 100
251 ATTATTCACACATCAGAAAGTATCATGTTATCGTGGCTGTAGAGCCAG 300
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuValAla 117
301 CTTTTCATATTTCTGGCCTTGGGGGCTAGTAATATCTCTCCGCGCT 350
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
351 CATGCTCTTGATAGAGTTTGTACTATTTTGGCCCTCCATTTACACAG 400
134 alIleMetHisGlnArgLeuGlyLeuGlnLeuAlaAlaSerTyrVal 150
401 TTATTCATGACACAGAGACTCGCTCCAGTTGGCAGCTGCATCTGGGTT 450
151 ThrGlyPheSerAsnSerValTirPLeuSerThrLeuThrLeuGlnLeuPr 167
451 ACTGGTTTGTAGTAACCTAGTGTGTTGTCTACACCTGACTCCAGCGCC 500
167 olauGlyAspProTyrValIleAspHisPheLeuGlyValProAla 184
501 ACCTGTGACCCCTATGTGTATGATATCATCTTCTGTGAGAGTCCCTGAC 550
184 euLeuLysLeuSerCysValGluThrThraAlaAsnGluAlaGluLeuPhe 200
551 TGTCTAAGTATCTTGTGTGTGAGACAACAGCAATGAGGCTGAACATTC 600
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSe 217
601 CTGTGAGTGAAGCTCTTCCATTAATACCCCTGACACTCATCTTATATC 650
217 rTyrAlaPheIleValAlaArgAlaValLeuArgIleGlnSerAlaGluGly 234
651 ATATGCTTTTATTTGTCGACACATATTGACGATACACTGCTGCAAGCTC 700
234 rGlnLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeu 250
701 GACAAAAGCATTTGGACATGTGTCCCATCTAATTTGGGTGTCCTT 750
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
751 TTTTATAGTACAGCCGCTCTGTGTACCTGCAACCACTTGGCCAGCTC 800
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaProm 284
801 CAAGGACCAAGAAAGATGTTCTCTCTCTTATGGAATATTGACACCA 850
284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
851 TGCTGATATCCCTTATATATACACTTAGAACAAGAGGTAAAGGAAGGC 900
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
|||||

```

```

901 TTTAAAGTTGGTGCACAGACTCTTCTTAATCAAGAA 939
seq_name: /SIDSz/gcgdata/geneseq/geneseqn/AA2001.DAT:AA508541
seq_documentation_block:
ID AA508541 standard; cDNA; 942 BP.
AC AA508541;
XX
XX
XX 26-SEP-2001 (first entry)
XX
XX
XX DNA encoding novel human G-protein coupled receptor (NGPCR).
XX
XX Human; novel G-protein coupled receptor; NGPCR; drug discovery;
XX diagnostic; ss.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 1..942
XX FT /*tag= a
XX FT /product= "Novel G-protein coupled receptor (NGPCR)."
XX
XX WO200151634-A1.
XX
XX 19-JUL-2001.
XX
XX 05-JAN-2001; 2001WO-US00589.
XX
XX 12-JAN-2000; 2000US-0175764.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX PA Walke DW, Wilganowski NL, Turner CA, Friedrich G, Abuin A;
XX PI Zambrowicz B, Sands AT;
XX DR WPI; 2001-442145/47.
XX DR P-PSDB; AAU04689.
XX
XX PT New human G protein-coupled receptor and polynucleotides encoding the
XX PT receptor, useful in identifying, selecting or validating new molecular
XX PT targets for drug discovery and in diagnostic or prognostic assays
XX
XX PS Claim 1; Page 61; 65pp; English.
XX
XX CC The sequence represents the coding sequence of novel human G-protein
XX CC coupled receptor (NGPCR). The NGPCR nucleic acid and protein are useful
XX CC in identifying, selecting or validating novel molecular targets for drug
XX CC discovery, as well as in diagnostic or prognostic assays. These are also
XX CC useful in microarrays or other assay formats, for screening collections
XX CC of genetic material from patients who have a particular medical condition
XX CC or for identifying mutations associated with a particular disease.
XX
XX SQ Sequence 942 BP; 222 A; 228 C; 187 G; 305 T; 0 other;

```

alignment\_scores:                      Length:      313

                    Quality: 1601.00                      Gaps:      0

                    Ratio: 5.115                              Percent Identity: 100.000

Percent Similarity: 100.000                      Percent Identity: 100.000

alignment\_block:

US-09-755-017-2 x AA508541                      ..

Align seg 1/1 to: AA508541 from: 1 to: 942

```

1 MetAsnTyrValAsnAspSerIleIleGlnGluPheIleLeuLeuGlyPh 17
|||||
1 ATGAATGGGTAATGACACATCATACAGAGATTATCTCGCGGTTT 50
17 eSerAspArgProTirPLeuGluPheProLeuValValPheLeuIles 34
|||||
51 CTGAGATCGACCTTGCTGGAGTTTCCACCTCTTGCTCTCTGATTT 100
51 CTGAGATCGACCTTGCTGGAGTTTCCACCTCTTGCTCTCTGATTT 100

```

34	erYrYrThValThrLiIlePheGlySnsLeuThrLiIleIleValSerArg	50
101	CTTACACGTGACCATCTTTGGCATGTGACCAATATTCTAGTGTACGC	150
51	LeuAspThrLysLeuHisThrProMetLysPhePheLeuThrAsnLeuSe	67
151	CTGGACACCAAACTCTTACACCCCATGATATTTTCTTACCAATGTATC	200
67	rLeuLeuAspLeuGlySyrThrThrCysThrValProGlnMetLeuValA	84
201	ACTCGTGATCTTTGGTTACACACATCTTACAGTCCACCAATCTGTGTA	250
84	snLeuGlySerLiIleArgLysValLiIleSerLysArgLysCysValIa	100
251	ATTATTCAGCATGACGAAAGATATACGTTATGTGTGGCTGTGACCCAG	300
101	LeuPheLiIlePheLeuAlaLeuGlyAlaThrGlnThrLeuLeuValA	117
301	CTTTTCATATTTCTGGCCCTTGGGGGCTACGAAATATCTTCTCCGGCGG	350
117	IMetSerPheAspArgPheValAlaLiIleCysArgProLeuHisTyrSer	134
351	CATGCTCTTTATAGGTTTGTAGCTATTCTTCGGCCCTCCATTAATCTCA	400
134	allMetHisGILArgLeuGlySnsLeuGlnLeuValAlaAlaSerTProVal	150
401	TTATCATGTACACCAAGACTCTGCTCCAGTTGGCAGCGCATCTGGGTT	450
151	ThrGlyPheSerAsnSerValThrPheLeuSerThrLeuThrLeuGlnLeuPr	167
451	ACTGGTTTATTAAGTCTAGTGTGGTTGTCTACCCCTGACTCTCCAGCGCC	500
167	OLEuGlyAspProThrValLiIleAspHisPheLeuGlyGluValProAla	184
501	ACTGTGTACCCCATGTGATGATGATCACTTCTCTGTGAAGTCCCTGCAC	550
184	euleuLysLeuSerCysValGlnThrThrLysAsnGluAlaGluLeuPhe	200
551	TGCTCAAGTATCTGTGTGTGTGACGACCAACAGCAATGAGCTGAATATTC	600
201	LeuValSerGluLeuPheHisLeuLiIleProLeuThrLeuLiIleLeuIleSe	217
601	CTTGCATAGTAGCTCTTCCATCTAATACCCCTGACACTCATCCTTATATC	650
217	rTyrrAlaPheLiIleValArgAlaValLiIleAspGlnIleGlnSerAlaGlnGlyA	234
651	ATATGCTTTTATTTGTCCGAGCAGATTTGAGAGATACAGTGTGCGAAAGTTC	700
234	rgGlnLysAlaPheGlyThrCysGlySerHisLeuLiIleValValSerLeu	250
701	GACAAAAAGCATTTGGGAGCATGTGTGCCATCTAATTTGGTGTCTCTT	750
251	PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe	267
751	TTTTTATGTACAGCCGCTCTGTCTTACTCTCAACACACTTTCGCCCAAGCTC	800
267	rLysAspGlnGlyLysMetValSerLeuPheTyrGlyLiIleLiIleAlaProM	284
801	CAGAGACCAAGAAAGAGTGTCTCTCTCTTATGAGGATATTGTGACCCA	850
284	etLeuAsnProLeuLiIleTyrThrLeuArgAsnLysGluValLysGlnGly	300
851	TGCTGAATCCCTTATATATATACACTTTGGAACAAGAGATGAAGGAAGGC	900
301	PheLysArgLeuValAlaArgValPheLeuLiIleLysLys	313
901	TTTAAAGGTTGGTTCGACAGAGCTCTTATATCAAGAAA	939

seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AA508542
seq\_documentation\_block:
ID AA508542 standard: cDNA: 1488 BP.

```

XX  AAS08542:
XX
XX  26-SEP-2001 (first entry)
XX
XX  ORF encoding novel human G-protein coupled receptor (NGPCR).
XX
XX  Human; novel G-protein coupled receptor; NGPCR; drug discovery;
XX  diagnostic; ss.
XX
XX  Homo sapiens.
XX
XX  Key      Location/Qualifiers
XX  Key      198..1139
XX  CDS       /*tag= a
XX  FT        /product= "Novel G-protein coupled receptor (NGPCR)"
XX
XX  WO200151634-A1.
XX
XX  19-JUL-2001.
XX
XX  05-JAN-2001; 2001WO-US00589.
XX
XX  12-JAN-2000; 2000US-0175764.
XX
XX  (LEXI-) LEXICON GENETICS INC.
XX
XX  Walke DW, Wilganowski NL, Turner CA, Friedrich G, Abulin A;
XX  Zambrowicz B, Sands AT;
XX
XX  WPI; 2001-442145/47.
XX
XX  P-PSDB; AAU04689.
XX
XX  New human G protein-coupled receptor and polynucleotides encoding the
XX  receptor, useful in identifying, selecting or validating new molecular
XX  targets for drug discovery and in diagnostic or prognostic assays
XX
XX  Disclosure; Page 62-63; 65pp; English.
XX
XX
XX  The sequence represents the open reading frame (ORF) encoding novel human
XX  G-protein coupled receptor (NGPCR). The NGPCR nucleic acid and protein
XX  are useful in identifying, selecting or validating novel molecular
XX  targets for drug discovery, as well as in diagnostic or prognostic
XX  assays. These are also useful in microarrays or other assay formats, for
XX  screening collections of genetic material from patients who have a
XX  particular medical condition or for identifying mutations associated with
XX  a particular disease.
XX
XX  Sequence 1488 BP; 421 A; 333 C; 254 G; 477 T; 3 other;
XX
XX
XX  alignment_scores:
XX      Quality: 1601.00      Length: 313
XX      Ratio: 5.115          Gaps: 0
XX  Percent Similarity: 100.000  Percent Identity: 100.000
XX
XX  alignment_block:
XX  US-09-755-017-2 x AAS08542 ..
XX
XX  Align seg 1/1 to: AAS08542 from: 1 to: 1488
XX
XX  1 MetAsnTrpValAsnApSerIleIleGlnGluPheIleuLeuGlyPh 17
XX  |||||
XX  198 ATGAATTTGGGTTAAATGCACACATCATACAGAGATTATTCGCTGGGTTT 247
XX
XX  17 eSerrAspArgProTyrPleuGluPheProLeuLeuValValPheLeuIles 34
XX  |||||
XX  248 CTCAGATCGACCTTGGCTGGAGGTTTCCACTCCCTTGCTGTCCTTGTGATT 297
XX
XX  34 eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
XX  |||||
XX  298 CTTACACTGTGACCATCTTTGGCAATGTGACCATATTATTAGTGTACGGC 347

```

```

51 LeuAspThrIleuHisThrProMetTyrPhePheLeuThrAsnLeuSe 67
   |||
348 CTGGACACCAACTTCATACCCCATGATATTTTCTTACCAATATCATC 397
   |||
67 rLeuLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuValA 84
   |||
398 ACTCTGGATCTTGTTCACACCATGTACAGTCCCAAAATCTAGTAA 447
   |||
84 snLeuCysSerIleArgLysValIleSerTyrArgGlyCysValAlaGln 100
   |||
448 ATTATATGCACATCAGGAAGATATCATGTCGTGCTGTAGTACCCAG 497
   |||
101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaValA 117
   |||
498 CTTTTCATATTTCTGGCCCTGGGGGGCTACTGAATATCTTCTCCGGCCG 547
   |||
117 lMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
   |||
548 CATGTCTTGTGATAGGTTTGTACTATTTCTCGGCCCTCCATTAATCAG 597
   |||
134 allMetHisGlnArgLeuCysLeuGlnLeuAlaAlaIleAspTyrVal 150
   |||
598 TTATCATGCACACAGACATCTGCTCCAGTTGGCAGCCGATCCTGGGTT 647
   |||
151 ThrGlyPheSerAsnSerValIlePheLeuSerThrLeuThrLeuGlnLeuPr 167
   |||
648 ACGGTGTTTGTAGTAACTAGTGTGGTGTACCCCTGACTCTCCAGCTGCC 697
   |||
167 olLeuCysAspProTyrValIleAspHisPheLeuCysGluValAlaProAl 184
   |||
698 ACCTCTGAGACCCCTATCTGTATGATCATCTTCTCTGGAAGTCCCTGCAC 747
   |||
184 euleuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200
   |||
748 TGTCTCAAGTTATCTGTGTGAGACAAACACCAATAGGCTGACTATTC 797
   |||
201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSe 217
   |||
798 CTGTGTAGTACGCTCTTCATTAATACCCCTGACACTCATTCCTTATATC 847
   |||
217 rTyrAlaPheIleValAlaArgAlaValLeuAlaGlnIleGlnSerAlaGluGlyA 234
   |||
848 ATATGCTTTTATTTGTCGACAGCATATTGAGGATACAGTCTGCTGAAGTCTC 897
   |||
234 rGlnLysAlaPheGlyThrCysGlySerHisLeuIleValIleSerLeu 250
   |||
898 GACAAAAGCATTTGGGACATGTGTGTTCCCATCTAATTTGGTGTCTCTT 947
   |||
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
   |||
948 TTTTATATGTAAGCGCTCTGTGTACCTGCAACACCTTCGGCCAGCTC 997
   |||
267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaProm 284
   |||
998 CAAGGACCAAGAAAGATGTTCTCTCTATGAAATATATGCAACCA 1047
   |||
284 etLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
   |||
1048 TGTGTGATTCCTTATATATATACCTTATGACAAAGGAGGTAAAGGAAAGGC 1097
   |||
301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
   |||
1098 TTTAAAGGTTGGTTCAGAGCTCTTCTTATATCAAGAAA 1136
   |||
seq_name: /SID2/gcdata/geneseq/geneseqn/NA2001.DAT:AAH31669
seq_documentation_block:
ID AAH31669 standard; DNA: 1071 BP.
XX AAH31669;
AC
XX
XX 30-JUL-2001 (first entry)
XX

```

```

DE Human olfactory receptor polynucleotide, SEQ ID NO: 242.
XX
XX
KW Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation; ds.
OS Homo sapiens.
PN WO200127158-A2.
PD 19-APR-2001.
XX
XX 06-OCT-2000; 2000WO-US27582.
XX
XX 08-OCT-1999; 99US-0158615.
XX
XX 24-FEB-2000; 2000US-0184809.
XX
XX (DIGI-) DIGISCENTS.
XX (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX WPI, 2001-290713/30.
XX
XX New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX
XX Claim 8; Page 273-274; 1857pp; English.
XX
XX The present sequence is one of a number of isolated polynucleotides
XX which encode polypeptides involved in olfactory sensation. The
XX polynucleotides can be used in screening for olfactory agonists and
XX antagonists. The methods allow for the determination of primary
XX scents and the identification of the odour receptors used to detect
XX these primary scents. The methods also enable determination of
XX secondary scents and the identification of combinations of odour
XX receptors that are involved in detecting such secondary scents.
XX This enables the construction of a scent representation (also called
XX a scent fingerprint or scent profile), which may be used to re-create
XX and edit scents. Libraries of olfactory receptors are useful for
XX determining the interaction pattern of a composition with the receptors,
XX and can be used for determining differences in the olfactory faculties
XX of different individuals.
XX
XX Sequence 1071 BP; 277 A; 248 C; 201 G; 345 T; 0 other:
XX
XX
XX alignment_scores:
XX      Quality: 1319.00      Length: 314
XX      Ratio: 4.517          Gaps: 1
XX      Percent Similarity: 92.994      Percent Identity: 82.484
XX
XX alignment_block:
XX US-09-755-017-2 x AAH31669 ..
XX
XX Align seg 1/1 to: AAH31669 from: 1 to: 1071
XX
XX 1 MetAsnTyrValAsnAspSerIleIleGlnGluPheIleLeuGluLys 17
XX |||
XX 1 ATGAATTTGGGTAATAAGATTCACCAAGGAGATTCATCTGTAGTATT 50
XX |||
XX 17 eSerAspArgProTyrLeuGlnPheProLeuLeuValIlePheLeuIle 34
XX |||
XX 51 CTCAGATCAACCATGCTGAGATTCACCCCTTGTGATGTTCTCTTTT 100
XX |||
XX 34 eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
XX |||
XX 101 CTTATATCTTGACAAATCTTGGCAATCTGACATATATCTGTGTACAT 150
XX |||
XX 51 LeuAspThrLysLeuHisThrProMetTyrPhePheLeuThrAsnLeuSe 67
XX |||
XX 151 GTGATTTCAAACTCCACACCCCTATGTACTTTTCTTACCAATCTCTC 200
XX |||

```

[illegible]

XX scent profile; scent fingerprint; scent representation; ds.  
OS Homo sapiens.  
PN WO200127158-A2.  
XX 19-APR-2001.  
PD 06-OCT-2000; 2000MO-US27582.  
PE 08-OCT-1999; 99US-0158615.  
PR 24-FEB-2000; 2000US-0184809.  
XX  
XX (DIGI-) DIGISCENTS.  
PA (YEDA ) YEDA RES & DEV CO LTD.  
XX  
XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;  
PI WPI; 2001-290713/30.  
XX  
DR  
XX  
XX New polynucleotides which encode polypeptides involved in olfactory  
PT sensation for identifying olfactory agonists and antagonists -  
PT  
XX  
XX Claim 8; Page 508-509; 1857pp; English.  
PS

The present sequence is one of a number of isolated polynucleotides which encode polypeptides involved in olfactory sensation. The polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. This enables the construction of a scent representation (also called a scent fingerprint or scent profile), which may be used to re-create and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors and can be used for determining differences in the olfactory faculties of different individuals.

```

50  sequence 1071 BP, 277 A, 248 C, 201 G, 345 T, 0 other
      alignment_scores:
      .      quality: 1319.00      length: 314
      .      ratio: 4.517      Gaps: 1
      Percent Similarity: 92.994      Percent Identity: 82.484

```

alignment\_block:

Align seg 1/1 to: AAH32250 from: 1 to: 1071

```

1  M6TA6NT6SP6VAL6AA6SP6SE6RI6LI6GL6IG6IP6HE6LE6LE6U6GL6YP6 17
1  ATGA6TTTGGG6RAAT6AAAG6AGTGTCCAC6AGSAG6TTCA6TCTG6TAG6TTT 50
17  eSe6rAsp6Arg6Pro6T6Leu6Glu6IP6He6Pro6Leu6Val6Val6Ph6e6Leu6Leu6 34
1  CTCA6ATCA6ACAC6GATG6CTAGA6GAT6TCCAC6CTTGTGAT6GTCTG6TTT 100
34  eT6T6Th6Val6Thr6Ile6He6Gly6Asn6Leu6Thr6Ile6Ile6Val6Se6r6Arg6 50
101  CCTA6TATCTG6CAAT6CTTGTGCA6ATCTGCA6AAATAT6CTGTGTCACAT6 150
51  Leu6Asp6Thr6Ile6Asn6Ist6Thr6Pro6Me6Ly6r6Ph6e6He6Leu6Thr6Asn6Leu6Se6 67
151  GTGA6TTC6CAAC6ATCACA6CCCTAT6GACT6TTTTTTCT6TAGCAAT6CTCTC 200
67  rLeu6Asp6rLeu6Cys6Tyr6Thr6Thr6Cys6Th6Val6Pro6Glu6Me6Leu6Val6Ala6 84
201  ACTCTGTGAC6CTTGTCT6CTATACCACA6AGTAC6AGTCTTCCACAAT6TGCTGTAA 250

```

```

84 snleucysserillearglyvalillessertyargglycysvalalagl 100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 ACATATSCAACACCAGGAAGTATATGATGATGCTGCTGGGCCAG 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 leupheilepheleuvalaleuglyalathrglutyileuvalava 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 CTTTTCATTTTCTGGCCCTGGGGTCCACAGAAATGCTCTCTGGGCCGT 350
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 lmetserpheaspargphevalalalecysargproleuhtyrserv 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 CATGTGCTTGATAGATTGTAGCTATTTTCGGCCCTCTCATTTACTCAA 400
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 alilemethisglinargleucysleuglneulalalaleaasertprval 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 TTTATCATGACACAGAGGCTCTGCTTCAGTTGGCAGCTGCATCTGGATT 450
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 ThnGlyphesersnservaitrpleuserthrleuthrleuglneupr 167
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
451 AGTGGCTTTAGCAATTCAGATTACAGTCCACCTGGACACTTAAGATGCC 500
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 oleucysaspprotyrvalilleasphileucysgluvalprol 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 ACTGTGGTGCACAAAGAACTGCATCTCTCTGTGAAGTCCCTGCTC 550
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 euleuylsleusercysvalglutthrthralaasnglualagluleuphe 200
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
551 TGCTCAAGTTGCTCCTGTTGACACAAACACCAATGAGCTGAACATTC 600
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 leuvalserlucleuphehisleuileproleuthrleuileuilese 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
601 TTTATCATGCTGCTATTCCTTCAATAACCGTGACACTCATCTTAAATC 650
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 rtyrvalaphellevalargalavalaleuargileglseralaglul 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
651 GTATGCTTTATTTGTCACACAGCTGTTGACAAATCCACTGCTGAAAGTC 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
234 rglulysalaphneglythrCysglyserhisleuilevalalserleu 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 AACGAAAGGATTTGGGACATGGGCTCCATCTAATTTGGTGTGCTCT 750
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 pheutyserthrAlavalservaltyrleuglproproserproserse 267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
751 TTTATGATGACACTATCTCATGTACTCTCAACACACCTTCCACCACTC 800
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
267 rlyaspglilglysmetvalserleuphetrglylilelialaprom 284
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
801 CAAAGACCGGGGAAAGATGCTTCTCTCTCTGTGGATCATTCACCCCA 850
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
284 etleuasnproleuiletyrthrleuargasnlysgluvalylsglul 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
851 TGCTGAATCCCTTATATATACACTTAGCAAAAGAGGTAAAGGAAGCC 900
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 phelysargleuvalalaArgvalpheluleuileylslys 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
901 TTTAAAGGTTGGTTCGCAAGAGCTCTTAAATCAAGAAA 940
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
seq_name: /SID57/gcdata/geneseq/geneseqn/NA2001.DAT:AAH32304
seq_documentation_block:
ID AAH32304 standard; DNA; 939 BP.
AAH32304:
30-JUL-2001 (first entry)
Human olfactory receptor polynucleotide, SEQ ID NO: 877.
Human olfactory receptor; OR; primary scent determination;
secondary scent determination; polypeptide library; odour receptor;
scent profile; scent fingerprint; scent representation; ds.
Homo sapiens.
OS
XX

```

```

PN WO200127158-A2.
XX
PD 19-APR-2001.
XX
PF 06-OCT-2000; 2000WO-US27582.
XX
PR 08-OCT-1999; 99US-0158615.
XX
PR 24-FEB-2000; 2000US-0184809.
XX
PA (DIGI-) DIGISCENTS.
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX WPI: 2001-290713/30.
DR
XX
PT New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -
XX
PS Claim 8; Page 530; 1857pp; English.
XX
CC The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.
XX
SQ Sequence 939 BP; 232 A; 228 C; 185 G; 294 T; 0 other:
XX
alignment_scores:
Quality: 1138.00 Length: 313
Ratio: 4.138 Gaps: 0
Percent Similarity: 87.859 Percent Identity: 70.607
alignment_block:
US-09-755-017-2 x AAH32304 ..
Align seg 1/1 to: AAH32304 from: 1 to: 939
1 MetAsnTrpValAsnAspSerIleIleGlnGluPheIleuLeuGlyP 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 ATGAAATTTGGGAAATAGAGACTCCCAAAAGAGATTATCTACTTGGCTT 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 eSerAspArgProTrpLeuGluPheProleuLeuValalPheLeuIle 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 CTCAGATAGAGGCTTGGTACAAATGCCCCCTTTTGGTGTCTGTATAT 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 ertYrThrValThrIlePheGlyAsnLeuThrIleIleLeuValalSer 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 CATACACAAATCACCATATTTGGCAATGTGTCATCATGATGGTGCATT 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 LeuAspThrLysLeuHisThrPrometTyRPhelPheLeuThrAsnLeu 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 CTGGATCCCAAACTTCATCTCCCATGATATTTCTCTCATCTACTCTC 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 rleuLeuAspLeuCysTyRThrThrcysThrValProGlnMetLeuVal 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 CARCTTAGATCTGCTATACCACTACAGTCCCTCATATGTGTGTA 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 snleucysserillearglyvalillessertyargglycysvalalagl 100
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 ATATTGTTGCACAAAGAACCATCATGCTATGCTGCTGTGGCCAC 300
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

101 LeuphelllephleuAlaLeuGlyAlaThrGluTyrLeuLeuAla 117  
 117 MetSerPheAspArgPheValAlaIleCysArgProLeuHISYrServ 134  
 351 TATGTCCTTTGACATATGATGCTGTTGACAGACCCCTCCACTATGATG 400  
 134 allMetHISGlnArgLeuGlyGlnLeuAlaAlaSerTPVal 150  
 401 TCATCATGATTAATGCTTCTGCTAAGATGAGCCCTTCATGCTTC 450  
 151 ThrGlyPheSerAsnSerValTyrLeuSerThrLeuGlnLeuPr 167  
 451 ATTGGTTTGGCACTCACTGAGCTGAGTCTTCTTACCTTAACATGCG 500  
 167 OLeuCysAspProTyrValIleAspHisPheLeuGlyAlaProAla 184  
 501 ACGCTGTGTCACAGAGAGTGGACCACTTTTCTGTGAGTGCCGAC 550  
 184 euleuLysLeuSerCysValGluThrThrAlaGlnGluLeuPhe 200  
 551 TTTCTAAGTTGTCATGCTGCTGACACAAAGCCATTGAGGCTGACTTC 600  
 201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIle 217  
 601 TTTCTTAGTACTAATTTCTTCTAATCCAGTACATTTATCTCTATC 650  
 217 rTyrAlaPheIleValAlaValAlaLeuArgIleGlnSerAlaGluGly 234  
 651 CTATGCTTCATAGCTCAAGCAGTATTAAATCAAGTCAGACAGAGGAC 700  
 234 rGlnGlnAlaPheGlyThrGlySerHisLeuIleValValSerLeu 250  
 701 GGCAAAACATTTGGACATGGGTCCACATGATTTGGTGGTCCCTC 750  
 251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267  
 751 TTTTATGAAACAGCATTTATATGATCTTCAACACCTTCATCCAC 800  
 267 rLysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284  
 801 TTAGGCTGCTGGAAAGATGTTCCCTCTTATGCAATCATCATCA 850  
 284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGlu 300  
 851 TGTGTAACCTCCCTCATCTACAGCTTAGAAATRAAGATATGAGAG 900  
 301 PheLysArgLeuValAlaArgValPheLeuIleLysLys 313  
 901 TTCAGAGGCTGATGCCAAGAAATCTTTTCTGTAACAAA 939  
 seq\_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DMT:AACT7475  
 seq\_documentation\_block:  
 ID AACT7475 standard; cDNA; 1442 BP.  
 AC AACT7475;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF3030 polynucleotide sequence SEQ ID NO:6059.  
 XX  
 KW Human: open reading frame; ORFX; detection; cytosstatic; hepatotropic;  
 KW vulnary; antipsoiatric; antiparkinsonian; noctropic; neuroprotective;  
 KW anticonvulsant; osteoporotic; antidiabetic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
 KW antiallergic; antibacterial; antifungal; antirheumatic; antihypertoid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;

severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 bone damage; cartilage damage; antinflammatory disease; coagulation;  
 thrombosis; contraceptive; ss.  
 OS Homo sapiens.  
 XX  
 XX WO200058473-A2.  
 PD 05-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US08621.  
 PF  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shinkets RA, Leach M;  
 XX WPI; 2000-602362/57.  
 DR P-PSDB; AAB43266.  
 DR  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X.  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PT  
 PS Claim 5; Page 5245-5246; 5507pp; English.  
 XX  
 CC AACT4446 to AACT77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytosstatic; hepatotropic; vulnary;  
 CC antipsoiatric; antiparkinsonian; noctropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antihypertoid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC  
 XX Sequence 1442 BP; 378 A; 319 C; 268 G; 477 T; 0 other:  
 SO  
 alignment\_scores:  
 Quality: 1138.00 Length: 313  
 Ratio: 4.138 Gaps: 0  
 Percent Similarity: 87.859 Percent Identity: 70.607  
 alignment\_block:  
 .US-09-755-017-2 x AACT7475 ..  
 1. Align seg 1/1 to: AACT7475 from: 1 to: 1442  
 1 MetAsnTyrValAsnAspSerIleIleGlnGluPheIleLeuGlnLys 17  
 301 ATGAATTTGGAAATAGACAGCTCCCAAAAGATTTATCTACTTGGCTT 350  
 17 eSerAspArgProTyrLeuGlnPheProLeuValAlaPheLeuLys 34  
 351 CTCAGATAGGCTTGGCTCAAAATGCCCTTTTGTGTCCTGTATATAT 400

```

34  eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
   |||.....|
401  CATACACAAATCACCACATATTTGGCAATGTGCATCATGAGGTGTCATT 450
   |||
51  LeuAspThrLysLeuHisThrPrometTyrPhePheLeuThrAsnLeuSe 67
   |||
451  CTGGATCCCAAACTCATCTACTCCATGATATTTCTTCTCTCAATATCTCTC 500
   |||
67  fLeuLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuVala 84
   |||
501  CATCTTAGATCTCTGATATGTGCTGTTCACAGACCTCCATATGTGTGTA 550
   |||
84  snLeuCysSerIleArgLysValIleSerTyrArgLysValAlaGln 100
   |||
551  ATATGTGTTCCACAAACAAACCATCATGCTATGCTGGCTGTGGCCAC 600
   |||
101  LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVala 117
   |||
601  CTCATCATCTTCCGTGGCCCTAGGTGCTACAGAGTGTCTCTCTGGCTGT 650
   |||
117  MetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerV 134
   |||
651  TATGTCTTGTGACAGATATGTGCTGTTCACAGACCCCTCCACTATGTAG 700
   |||
134  allMetHisGlnArgLeuCysLeuGlnLeuAlaAlaIleSerTyrPval 150
   |||
701  TCATCATGATATTTGTTGTTCTGCTAAGATGACAGCTTCTCATGTGCTC 750
   |||
151  ThrGlyPheSerAsnSerValTyrPleuSerThrLeuThrLeuGlnLeuP 167
   |||
751  ATTGGTTTGGCAACTCAGTGTGCTGCTCTTCTGCTCATCTTAACTGTC 800
   |||
167  oleuCysAspProTyrValIleAspHisPheLeuCysGluValProAla 184
   |||
801  ACCCTGTGTCACACAGAGTGGACACCTTTTCTGTGAGTCCCTGCAC 850
   |||
184  euLeuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPhe 200
   |||
851  TTCTCAAGTTGTCATGTGTCGACACAAAGCTATTGAGGTGAGCTTCT 900
   |||
201  LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSe 217
   |||
901  TTCTTAGTCTACTAATTTCTTAATTTCCAGTACATTTGATCTCTCATCTC 950
   |||
217  fTyrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyA 234
   |||
951  CTATGGCTTCATAGCTCAACAGCTATTTAAATCAGTCCAGCAGAGAGAC 1000
   |||
234  rglInLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 250
   |||
1001  GGCACAAAGCATTTGGGACATGTGGGTCCACATGATGTGCTGCTCCTC 1050
   |||
251  PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSe 267
   |||
1051  TTTTATGGAAACAGCCATTATATATCTTCAACACCATTCATCCACCTTC 1100
   |||
267  fLysAspGlnGlyMetValSerLeuPheTyrGlyIleIleAlaProM 284
   |||
1101  TAAAGACTGGGAAAGATGGTTTCCCTTCTTATGGAATCATCATCAATCCA 1150
   |||
284  eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
   |||
1151  TGTGTAAGTCTCCATCTACCTACAGCCTTAAGAAATATGAAGAGAGGCC 1200
   |||
301  PheLysArgLeuValAlaArgValPheLeuIleLysLys 313
   |||
1201  TTCAAGAGGCTGATGCCAAGAAATCTTTTCTGTAAGAAA 1239
   |||
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH32231
seq_documentation_block:
ID  AAH32231 standard: DNA: 648 BP.

```

```

XX  AAH32231;
AC  30-JUL-2001 (first entry)
XX  30-JUL-2001 (first entry)
XX  Human olfactory receptor polynucleotide, SEQ ID NO: 804.
DE  Human: olfactory receptor; OR: primary scent determination;
KW  secondary scent determination; polypeptide library; odour receptor;
KW  scent profile; scent fingerprint; scent representation; ds.
XX  Homo sapiens.
OS  WO200127158-A2.
XX  19-APR-2001.
XX  06-OCT-2000; 2000WO-US27582.
XX  08-OCT-1999; 99US-0158615.
XX  24-FEB-2000; 2000US-0184809.
XX  (DIGI-) DIGISCENTS.
XX  (YEDA ) YEDA RES & DEV CO LTD.
XX  Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX  WPI: 2001-290713/30.
XX  New polynucleotides which encode polypeptides involved in olfactory
XX  sensation for identifying olfactory agonists and antagonists -
XX  Claim 8; Page 502; 1857pp; English.
XX  The present sequence is one of a number of isolated polynucleotides
XX  which encode polypeptides involved in olfactory sensation. The
XX  polynucleotides can be used in screening for olfactory agonists and
XX  antagonists. The methods allow for the determination of primary
XX  scents and the identification of the odour receptors used to detect
XX  these primary scents. The methods also enable determination of
XX  secondary scents and the identification of combinations of odour
XX  receptors that are involved in detecting such secondary scents.
XX  This enables the construction of a scent representation (also called
XX  a scent fingerprint or scent profile), which may be used to re-create
XX  and edit scents. Libraries of olfactory receptors are useful for
XX  determining the interaction pattern of a composition with the receptors,
XX  and can be used for determining differences in the olfactory faculties
XX  of different individuals.
XX  Sequence 648 BP; 144 A; 166 C; 132 G; 206 T; 0 other:

Alignment_scores:
Quality: 1083.00 Length: 216
Ratio: 5.061 Gaps: 0
Percent Similarity: 99.074 Percent Identity: 99.074

alignment_block:
US-09-755-017-2 x AAH32231 ..
Align seg 1/1 to: AAH32231 from: 1 to: 648

68  LeuLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuValas 84
   |||
1  CTCCTGGATCTTGTTCACACACATGTACAGTCCCAAAATGCTACTATAA 50
   |||
84  nLeuCysSerIleArgLysValIleSerTyrArgLysValAlaGln 101
   |||
51  TTTATCAGACATCAGCAAGTATATCATGCTATGCTGTGATGCCAC 100
   |||
101  euPheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuAlaVal 117
   |||
101  TTTTCATATTTTGGCTTGGGGCTACAGAAATATCTTCTCGCGCCGTC 150
   |||

```



```

118 MetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerVa 134
151 ATGTCCTTTGATAGGTTTGTACCTATTTCGCGCTCTCATTTACTCACT 200
134 IleMetHisGlnArgLeuCysLeuGlnLeuAlaIleAspTyrPValT 151
201 TATCATGCACAGACAGCTGCTCCAGTTGGACGTCATCCAGGTTA 250
151 hnglyPheSerAsnSerValTyrPLeuSerThrLeuThrLeuGlnLeuPro 167
251 CCGGTTTAGTAACTCAGAGTGGTGTCTACCCCTGACTCTCCAGCTGCCA 300
168 LeuCysAspProTyrValIleAspHisPheLeuCysGluValProAlaLe 184
301 CTCTGTGACCCCTATGTGTATAGACCACTTCTCTGTGAAGTCCCTGCACT 350
184 uLeuLysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPheL 201
351 GCTCAAGTTATCTTGTGTGAGACAACAGCAATAGAGCTGAACCTATTCC 400
201 euValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSer 217
401 TTGTGAGTAGGCTCTCCATTAATACCCCTGACACTCATCTTATATCA 450
218 TyrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluIle 234
451 TATGCTTTATTTGTCGAGACAGTATTTGAGGATACAGTCTGCTGAAGGTG 500
234 gGlnLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeuP 251
501 ACAAAAAGCATTTGGGACATGTGGTCCCATCTAATTTGGTGTCTCTTT 550
251 heTyrSerThrAlaValSerValTyrLeuGlnProProSerProSerSer 267
551 TTAATAGTACAGCCGCTCTGTGTACTGCAACACACCTTCGCCACCTCC 600
268 LysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 283
601 AAGGACCAAGGAAGATGTTCTCTCTTATGGAATCATTTGCACCC 648
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH32025
seq_documentation_block:
ID AAH32025 standard; DNA; 979 BP.
XX
AC AAH32025:
XX
DT 30-JUL-2001 (first entry)
XX
DE Human olfactory receptor polynucleotide, SEQ ID NO: 598.
XX
KW Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation; ds.
XX
OS Homo sapiens.
XX
PN WO200127158-A2.
XX
PD 19-APR-2001.
XX
PF 06-OCT-2000; 2000WO-US27582.
XX
PR 08-OCT-1999; 99US-0158615.
PR 24-FEB-2000; 2000US-0184809.
XX
PA (DIGI-) DIGISCENS.
PA (YEDA) YEDA RES & DEV CO LTD.
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX WPI; 2001-290713/30.

```

```

XX
PT New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -
XX
PS Claim 8; Page 420; 1857pp; English.
XX
CC The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.
XX
SQ Sequence 979 BP; 225 A; 260 C; 208 G; 286 T; 0 other:

alignment_scores:
Quality: 954.00 Length: 309
Ratio: 3.614 Gaps: 2
Percent Similarity: 85.437 Percent Identity: 60.518

alignment_block:
US-09-755-017-2 x AAH32025
Align seg 1/1 to: AAH32025 from: 1 to: 979
1 MetAsnTyrValAsnAspSerIleIleGlnGluPheIleLeuGluLypH 17
1 ATGGAAGAAGCATTAACAGACAGACCTTCTGTGATTCATCTCTCGGGCT 50
17 eSerAspArgProTyrPLeuGlnPheProLeuValIlePheLeuIle 34
51 CTCCAAGAGCCCTCAGTGGAAACAGCTCTTTGGTCATCTTATCA 100
34 eTyrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50
101 TCTACTTTCTGAGCTTCTGTGGCAATGCAACCATTAATCACTTATTA 150
51 LeuAspThrLysLeuHisThrPrometTyrPhePheLeuThrAsnLeu 67
151 GTAGATCCGCGCTCCATACCCCTATGATTTCTCTCTCCATCTCTC 200
67 rLeuLeuAspPLeuCysTyrThrThrCysThrValProGlnMetLeuVal 84
201 TTTTATGATCTTTGTTGACCACTGTACTGCTCCAGACACCTGTGCA 250
84 snLeuCysSerIleArgLysValIleSerTyrArgGlyCysValAlaGln 100
251 ACTTTAAGGGGAAGACAGACCATCATCACTATGCTGCTGACCCAG 300
101 LeuPheIlePheLeuAlaLeuGlyAlaIleGluTyrLeuLeuLeuAla 117
301 CTATTATTTGCTTTGGGAGCTCGGGGGAGTGGAGTGTCTTATTTCTG 350
117 aMetSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSer 133
351 CCATGCGCTATGACCGCTATGACCGCTGCGCCACCTCCATCAATG 400
134 ValIleMetHisGlnArgLeuCysLeuGlnLeuAlaIleAspTyrPVal 150
401 GTAGCATGATCCCACTTGTGTGAGAGTGTGTATCACTGTGGCT 450
150 lThnglyPheSerAsnSerValTyrPLeuSerThrLeuThrLeuGlnLeuP 167
451 CACAGGTTTGGCAATTTGTGTATACAGACAGCATTTGACCATGCTCTC 500

```

```

167 roleuAspProtyrValIleAspHisPheLeuGluValProAla 183
    |||||
501 CCCCTGTGATMAAACCAAGTGATGATTTCTTGTAGATCCAGTG 550
    |||||
184 leuLeuValSerCysValIgluThrAlaAsnIuAgluLeuPh 200
    |||||
551 ATGCTGAACCTGCTGCACCAACACCTCATCAAGAGGCTGAATCTT 600
    |||||
200 eleuValSerCyluLeuPheHisLeuIleProLeuThrLeuIle 217
    |||||
601 TCGTGTAGTGTCTTCTTGTGGGCGCTCTCTCATCATCTTACCAT 650
    |||||
217 eTyThrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGlu 233
    |||||
651 CCAATGCTACATTTACTCATGCACTCTCTGAAGATMAAGTCAAGCTCA 700
    |||||
234 ArgGluLysAlaPheGlyThrCysGlySerHisLeuIleValSer 250
    |||||
701 AGGCACAGAGCTTTGGAACCTGGTTCACCTCCGTGAGTATCAT 750
    |||||
250 uPheTySerThrAlaValSerValTyThrLeuGlnProProSerPro 267
    |||||
751 TTTCTTTGGACACTCATCTTCATGTACCTCAGCCTCCCTCAGTTATT 800
    |||||
267 eTyAspArgGlnGlyLysMetValSerLeuPheTyThrGlyIleIleAlaPro 283
    |||||
801 CACAGGATGTGAACAAAGCATGTGCACTCTTCTATACTGTGACCTCT 850
    |||||
284 MetLeuAsnProLeuIleTyThrLeuArgAsnLysGluValLysGlu 300
    |||||
851 CTACTGATCCCTAATTACTCTGAGAACAGAGAGTCAAAAGGGGC 900
    |||||
300 yPheLys.ArgLeuValAlaArg 307
    |||||
901 AACTAGAGAGACTAGTGGGAGG 923
seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:AAH32067
seq_documentation_block:
ID AAH32067 standard; DNA; 927 BP.
XX AAH32067;
AC
XX
XX
DT 30-JUL-2001 (first entry)
XX
DE Human olfactory receptor polynucleotide, SEQ ID NO: 640.
XX
XX Human; olfactory receptor; OR; primary scent determination;
XX secondary scent determination; polypeptide library; odour receptor;
XX scent profile; scent fingerprint; scent representation; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200127158-A2.
XX
XX 19-APR-2001.
XX
XX 06-OCT-2000; 2000MO-US27582.
XX
XX 08-OCT-1999; 99US-0158615.
XX
XX 24-FEB-2000; 2000US-0184809.
XX
XX (DIGI-) DIGISCENTS.
XX (YEDA) YEDA RES & DEV CO LTD.
XX
XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX WPI; 2001-290713/30.
XX
XX New polynucleotides which encode polypeptides involved in olfactory
XX sensation for identifying olfactory agonists and antagonists -
XX

```

```

PS Claim 8: Page 437; 1857pp; English.
XX
CC The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.
XX
XX Sequence 927 BP; 209 A; 253 C; 176 G; 289 T; 0 other;
XX
XX
XX alignment_scores:
XX Quality: 944.00 Length: 302
XX Ratio: 3.688 Gaps: 0
XX Percent Similarity: 84.768 Percent Identity: 59.603
XX
XX alignment_block:
XX us-09-755-017-2 x AAH32067
XX
XX Align seg 1/1 to: AAH32067 from: 1 to: 927
XX
XX 5 AsnAspSerIleIleGlnIuPheIleLeuLeuGlyPheSerAspArg 21
XX |||||
XX 13 AATGAGAGTCCCTAATGATTCATCTCTGAGCTCTCAACACCACC 62
XX |||||
XX 21 oTyThrLeuGluPheProLeuValAlaPheLeuIleSerTyThrVal 38
XX |||||
XX 63 TCGCTGTGAGAGCTCTCTCTTGTATTTGCTCTCTCTCTCTCTCTGA 112
XX |||||
XX 38 hTlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLys 54
XX |||||
XX 113 CCCTGTGGGAACCTTCACCATATCATCATATCTATCTGATGCCCTCT 162
XX |||||
XX 55 LeuHisThrProMetTyThrPhePheLeuThrAsnLeuSerLeuLeuAsp 71
XX |||||
XX 163 CTTCATACCCCAATGATGATTTTTCAGACAACTCTCTTACTGAGCAT 212
XX |||||
XX 71 uCysTyThrThrCysThrValProGlnMetLeuValAsnLeuCysSer 88
XX |||||
XX 213 CTGCTTCACTACACTAGCCCTGCTCTCAGACCTTGAATTAATCTGCA 262
XX |||||
XX 88 leuArgLysValIleSerTyArgGlyCysValAlaGlnLeuPheIlePhe 104
XX |||||
XX 263 CAAGAAGAGGATCACTTACGGGTGTGTGTGGCGCACTCTATATTTCT 312
XX |||||
XX 105 leuAlaLeuGlyAlaThrGluTyThrLeuLeuAlaValMetSerPheAs 121
XX |||||
XX 313 CTGGCACTGGGCTGCCAOTGAATGATTCCTGTGGACATGGCCTTGGA 362
XX |||||
XX 121 pArgPheValAlaIleCysArgProLeuHisTySerValIleMetHisG 138
XX |||||
XX 363 TCGGTACATGTGTCTGCAACCCCTCCACTGTATGATCATCATCAAGC 412
XX |||||
XX 138 lnaArgLeuCysLeuGlnLeuAlaAlaIleSerTyThrValThrGlyPheSer 154
XX |||||
XX 413 CACGGCTTTGCCAACAAGCTGGCATGTATCTCGGGCTCAAGTGGTGGCT 462
XX |||||
XX 155 AsnSerValTyThrPheSerThrLeuThrLeuGlnLeuProLeuCysAsp 171
XX |||||
XX 463 AGTTCCCTAATGCATGCAACTTTTACTCTGCAATTCGCTCTGTGGCAA 512
XX |||||
XX 171 oTyThrValIleAspHisPheLeuCysGluValAlaProAlaLeuLeuLysLeu 188
XX |||||
XX 513 CCATAGCGCTGCACTTTTATTTGCGAAGATACAGCTCTTCAAGTTGG 562

```

```

188 erCysValGluThrThrAlaAsnGluAlaGluLeuPheLeuValSerGlu 204
    ::::::::::::::::::::|
563 CTGGTGTGACACACACGTGCATGAATGTCCTTTCTTGTAGTGT 612
    ::::::::::::::::::::|
205 LeuPheHisLeuLeuProLeuThrLeuLeuLeuLeuSerTyraAlaPhe11 221
    ::::::::::::::::::::|
613 CGTTGTTGTCATCCACACACACTGCATCTCCATCTCTATAGGCTTCA 662
    ::::::::::::::::::::|
221 eValArgAlaValLeuArgGluLeuSerAlaGluGlyArgGlnValAlaP 238
    ::::::::::::::::::::|
663 AACCTAACCTGTGTGAGAGATCAATCAATAGAGCAAGCAAGACCT 712
    ::::::::::::::::::::|
238 heGlyThrCysGlySerHisLeuLeuValSerLeuPheLeuSerThr 254
    ::::::::::::::::::::|
713 TCAGCACTGCTCTCTCCACCTTACAGTGTGATATATCTATATGACAC 762
    ::::::::::::::::::::|
255 AlaValSerValTyrrLeuGluProProSerProSerSerLysAspGln 271
    ::::::::::::::::::::|
763 ATAATCTAGTGTACCTGCACCTAGTACACCTATATATATGACAC 812
    ::::::::::::::::::::|
271 yLysMetValSerLeuPheTyrrGlyLeuLeuAlaProMetLeuAsnPro 288
    ::::::::::::::::::::|
813 GAAGTTATCTCCCTTCTACACCATGCTGACCCCACTTAAATCCNA 862
    ::::::::::::::::::::|
288 euLeuTyrrThrLeuArgAsnLysGluValLysGluGlyPheLysArgLeu 304
    ::::::::::::::::::::|
863 TCATCTATATCTTAAAGAACAAAGATATGAAAGAGGCTCTGAGGAAC 912
    ::::::::::::::::::::|
305 ValAla 306
    ::::::::::::::::::::|
913 CTCTCG 918

```

seq\_name: /SIS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH32308

seq\_documentation\_block:

ID AAH32308 standard; DNA; 933 BP.

AC AAH32308;

DT 30-JUL-2001 (first entry)

DE Human olfactory receptor polynucleotide, seq ID NO: 881.

KW Human; olfactory receptor; OR; primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

OS Homo sapiens.

PN WO200127158-A2.

PD 19-APR-2001.

PF 06-OCT-2000; 2000WO-US27582.

PR 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

PA (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

DR WPI: 2001-290713/30.

PT New polynucleotides which encode polypeptides involved in olfactory

PS sensation for identifying olfactory agonists and antagonists -

CC Claim 8; Page 531-532; 1857pp; English.

CC The present sequence is one of a number of isolated polynucleotides

CC which encode polypeptides involved in olfactory sensation. The

polynucleotides can be used in screening for olfactory agonists and antagonists. The methods allow for the determination of primary scents and the identification of the odour receptors used to detect these primary scents. The methods also enable determination of secondary scents and the identification of combinations of odour receptors that are involved in detecting such secondary scents. This enables the construction of a scent representation (also called a scent fingerprint or scent profile), which may be used to re-create and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors, and can be used for determining differences in the olfactory faculties of different individuals.

Sequence 933 BP; 205 A; 239 C; 187 G; 302 T; 0 other;

alignment\_scores:

Quality:	941.00	Length:	302
Ratio:	3.734	Gaps:	0
Percent Similarity:	83.444	Percent Identity:	60.265

alignment\_block:

US-09-755-017-2 x AAH32308 ..

Align seg 1/1 to: AAH32308 From: 1 to: 933

```

4 ValAsnAspSerLeuLeuGluPheLeuLeuGlyPheSerAsp 20
    ::::::::::::::::::::|
19 GTCATATGCTAGCTCTGAGGGGTACTTATTTATGTTGAGATTTCTAATG 68
    ::::::::::::::::::::|
20 gProTrPLeuGluPheProLeuValValPheLeuLeuSerTyrrThy 37
    ::::::::::::::::::::|
69 GCCTCATCTGGAAGTAGTATCTTGTGCTGCTGATCTCTACTTGA 118
    ::::::::::::::::::::|
37 aLThrLeuPheGlyAsnLeuThrLeuLeuValSerArgLeuAspThr 53
    ::::::::::::::::::::|
119 TGACACTGATAGGAACCTGTTTCATCATCATCTGTCATACCTGAGCTC 168
    ::::::::::::::::::::|
54 LysLeuHisThrProMetTyrrPhePheLeuThrAsnLeuSerLeuAsn 70
    ::::::::::::::::::::|
169 CATCTGCACACACCAATGATCTTCTTCAACCTTCATTTCTGGA 218
    ::::::::::::::::::::|
70 pleuGlyTyrrThrCysThrValProGlnMetLeuValAsnLeuCys 87
    ::::::::::::::::::::|
219 TCTCTCTACACACACCTCTATCCCTCAGTGTGCTGATCTCTGG 268
    ::::::::::::::::::::|
87 eLLeuArgLysValLeuSerTyrrArgGlyCysValAlaGlnLeuPhe 103
    ::::::::::::::::::::|
269 GCCCGGAAAAAGACCATCTTATGCTGTGTCATGATTCACCTTACTT 318
    ::::::::::::::::::::|
104 PheLeuAlaLeuGlyAlaThrGluTyrrLeuLeuAlaValMetSerP 120
    ::::::::::::::::::::|
319 GTTCTGCGACCTGGGAACACAGAGTGTCTCTGCTGCTGATGCTGA 368
    ::::::::::::::::::::|
120 eaSPARgPheValAlaLeuCysArgPProLeuHisTyrrSerValIle 137
    ::::::::::::::::::::|
369 TGAACGTTATGACAGCTGTGTGAGACCTTGTGCTTACAGCTCTCAT 418
    ::::::::::::::::::::|
137 isGlnArgLeuGlyLeuGlnLeuAlaAlaLeuSerTyrrValThrG 153
    ::::::::::::::::::::|
419 ACCCTGTTTCTGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 468
    ::::::::::::::::::::|
154 SerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeu 170
    ::::::::::::::::::::|
469 ACCAAGCTCAGCACTTCTCTCTTCTTCTGCTGCTGCTGCTGCTG 518
    ::::::::::::::::::::|
170 pProTyrrValIleAspHisPheLeuGlyValProAlaLeuLeuLys 187
    ::::::::::::::::::::|
519 ACACCGGCAAGTAGATCATTCTTCTGTGAGTGTGACGACCTCTG 568
    ::::::::::::::::::::|
187 euSerCysValGluThrThrAlaAsnGluAlaGluLeuPheLeuVal 203
    ::::::::::::::::::::|
569 TATGCTGTGTGATACCATGATCAATGAGCTGACCTCATGATCACA 618
    ::::::::::::::::::::|

```

204 GluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSerTyraLaph 220  
 619 TCATATTTTCTCTCACTCTCATCCCTATCTCTCTTATATGATGC 668  
 220 eileValaArgAlaValLeuArgIleGlnSerAlaGluGlyArgGlnLysA 237  
 669 CATGCGCGAGCTGACTGAGATGACGACCACTGGCTTCAGAAAG 718  
 237 IapheGlyThrCysGlySerHisLeuIleValSerLeuPheTySer 253  
 719 TGTTTGAACTGTGAGCTCATCTATGGCTATCTCTCTTTTCATT 768  
 254 ThrAlaValSerValTyLeuGlnProProSerProSerLysAsp1 270  
 769 CCGCGCATGTCATATATCTCCAGCCACCTCAGAAATCTCAAGATCA 818  
 270 nglyLysMetValSerLeuPheTyrglyIleIleAlaProMetLeuAsp 287  
 819 AGCGAAGTTCATTCGCCCTTTTATCTGTCACACCTAGCTTAAAC 868  
 287 rleuLeuIleTyThrLeuArgAsnLysGluValLysGluGlyPheLysArg 303  
 869 CTCTAATCTACACCTCAGAAACAGTTGTAGAGGGGCGCATGAGAGA 918  
 304 LeuVal 305  
 919 CTATAG 924  
 seq\_name: /SID2/gcgcdata/geneseq/geneseq/NA2001.DAT:AAH31645  
 seq\_documentation\_block:  
 ID AAH31645 standard; DNA; 936 BP.  
 AC AAH31645;  
 DT 30-JUL-2001 (first entry)  
 DE Human olfactory receptor polynucleotide, SEQ ID NO: 218.  
 XX  
 KW Human; olfactory receptor; OR; primary scent determination;  
 KW secondary scent determination; polypeptide library; odour receptor;  
 KW scent profile; scent fingerprint; scent representation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200127158-A2.  
 PD 19-APR-2001.  
 XX  
 PF 06-OCT-2000; 2000WO-US27582.  
 XX  
 PR 08-OCT-1999; 99US-0158615.  
 PR 24-FEB-2000; 2000US-0184809.  
 XX  
 PA (DIGI-) DIGISCENTS.  
 PA (YEDA) YEDA RES & DEV CO LTD.  
 XX  
 PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;  
 DR WPI: 2001-290713/30.  
 XX  
 PT New polynucleotides which encode polypeptides involved in olfactory  
 PT sensation for identifying olfactory agonists and antagonists -  
 XX  
 PS Claim 8: Page 263; 1857pp; English.  
 CC The present sequence is one of a number of isolated polynucleotides  
 CC which encode polypeptides involved in olfactory sensation. The  
 CC polynucleotides can be used in screening for olfactory agonists and  
 CC antagonists. The methods allow for the determination of primary  
 CC scents and the identification of the odour receptors used to detect  
 CC these primary scents. The methods also enable determination of

secondary scents and the identification of combinations of odour  
 CC receptors that are involved in detecting such secondary scents.  
 CC This enables the construction of a scent representation (also called  
 CC a scent fingerprint or scent profile), which may be used to re-create  
 CC and edit scents. Libraries of olfactory receptors are useful for  
 CC determining the interaction pattern of a composition with the receptors,  
 CC and can be used for determining differences in the olfactory faculties  
 CC of different individuals.  
 XX  
 SQ Sequence 936 BP; 222 A; 226 C; 188 G; 300 T; 0 other;  
 alignment\_scores:  
 Quality: 939.00 Length: 305  
 Ratio: 3.711 Gaps: 0  
 Percent Similarity: 82.951 Percent Identity: 60.656  
 alignment\_block:  
 US-09-755-017-2 x AAH31645 ..  
 Align seg 1/1 to: AAH31645 from: 1 to: 936  
 1 MetaAsnTrpValAsnAspSerIleIleGlnGluPheIleLeuGlyph 17  
 1 ATGATCAGAAATAATGGAAGTCTTCTACAGTGAATTAATCTACTAGGTTT 50  
 17 eSerAspArgProThrLeuGluPheProLeuValValPheLeuIle 34  
 51 CTCTGACAGGCTTCAGCTGAGCTAGCTCTTGTGGTCTTTGATCT 100  
 34 eTyThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArg 50  
 101 TCTATATCTTCACTTCTGCTGGGACAAACCATCATTTATATATCTCAC 150  
 51 LeuAspThrLysLeuHisThrProMetTyrrPhePheLeuThrAsnLeu 67  
 151 TTGACCCACATCTTCACACTCTATGATATTTTCTTCCAACTAAG 200  
 67 rleuLeuAspLeuGlyTyThrThrCysThrValProGlnMetLeuVal 84  
 201 CTTTGTGATCTGTGTACCAACCGCATGTGTCCACACCTCTGGTTA 250  
 84 snLeuGlySerIleArgLysValIleSerTyrrArgLysValAlaGln 100  
 251 ATCTCAGGGGACGACGACAAATCAATCTCTATGTGCTGTGATTCAG 300  
 101 LeuPheIlePheLeuAlaLeuGlyAlaThrGluTyrrLeuLeuAlaVal 117  
 301 CTGTACATCTCTCTAGGCTGGGATCTACAGAAATGCTCTCTTAGAGT 350  
 117 lmetSerPheAspArgPheValAlaIleCysArgProLeuHisTySer 134  
 351 GATGTATTTGACCGCTATGACAGCTGTTCAGCCCTCCACTACACAG 400  
 134 alileMetHisGlnArgLysLeuGlnLeuIleAlaIleSerTyrrVal 150  
 401 TATCATGACCTTGTCTGTATGCTGATGCTGATGCTTACTTATGGTTC 450  
 151 ThrGlyPheSerAsnSerValThrLeuSerThrLeuThrLeuGlnLeuP 167  
 451 ATTGCTTTTGCCAAATCCATATTCAGAGCGGCTCTCATCTGCTTTTAA 500  
 167 oleuGlyAspProTyrrValIleAspHisPheLeuGlyAlaProAla 184  
 501 ACTTGTGAGAAATAATTAATTAAGACCTTCTTGTGAGTCTTCAT 550  
 184 euleuLysLeuSerCysValGluThrThrAlaAsnGluLeuPhe 200  
 551 TGCTCAAGCTTGCTGTGTGACACTACTATGATATTAATCTGAACTTTC 600  
 201 LeuValSerGluLeuPheHisLeuIleProLeuThrLeuIleLeuIle 217  
 601 TTGTGAGTGTCTATTTCTTCTTGTACCTGTTGCAATATATATCTTC 650

```

217 rTyrAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyA 234
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
651 CTrATGTCAGATTGTCAGGGCAGTCATGAGTAAAGTTAGCAACAGGGC 700
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 rGlnLysAlaPheGlyThrCysGlySerHisLeuIleValSerLeu 250
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
701 AGAGAAAGTGTGGACATGTGGCTCCACCTCACAGTGGTTCCCTG 750
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 PheTyrSerThrAlaValSerValTyrLeuGlnProProSerProSer 267
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
751 TTCTACGGGACAGCTATCTATCTTACCCGACCGGCAACACTACTC 800
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 rTysAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaPro 284
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 TCAGATCAGGGCAGGCAAGTCTCTCTCTTCTACACCATCATACCCCA 850
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
284 eLeuAsnProLeuIleTyrThrLeuArgAsnLysGluValLysGluGly 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 TGATCAACCCCTCATATATACACTGAGGACAGAGATGTGAAGAGAGCA 900
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 PheLysArgLeuVal 305
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 CTTAAGAAAGTCTCTC 915
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH32306

```

```

seq_documentation_block:
ID   AAH32306 standard; DNA; 1011 BP.
XX
XX   AAH32306;
XX
XX   30-JUL-2001 (first entry)
XX
XX   Human olfactory receptor polynucleotide. SEQ ID NO: 879.
XX
XX   Human; olfactory receptor; OR; primary scent determination;
XX   secondary scent determination; polypeptide library; odour receptor;
XX   scent profile; scent fingerprint; scent representation; ds.
XX
XX   Homo sapiens.
XX
XX   MO200127158-A2.
XX
XX   19-APR-2001.
XX
XX   06-OCT-2000; 2000MO-US27582.
XX
XX   08-OCT-1999; 99US-0158615.
XX
XX   24-FEB-2000; 2000US-0184809.
XX
XX   (DIGIT-) DIGISCENTS.
XX   (YEDA ) YEDA RES & DEV CO LTD.
XX
XX   Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX   WPI: 2001-290713/30.
XX
XX   New polynucleotides which encode polypeptides involved in olfactory
XX   sensation for identifying olfactory agonists and antagonists -
XX
XX   Claim 8; Page 531; 1857pp; English.
XX
XX   The present sequence is one of a number of isolated polynucleotides
XX   which encode polypeptides involved in olfactory sensation. The
XX   polynucleotides can be used in screening for olfactory agonists and
XX   antagonists. The methods allow for the determination of primary
XX   scents and the identification of the odour receptors used to detect
XX   these primary scents. The methods also enable determination of
XX   secondary scents and the identification of combinations of odour
XX   receptors that are involved in detecting such secondary scents.
XX   This enables the construction of a scent representation (also called
XX   a scent fingerprint or scent profile), which may be used to re-create

```

and edit scents. Libraries of olfactory receptors are useful for determining the interaction pattern of a composition with the receptors, CC and can be used for determining differences in the olfactory faculties CC of different individuals.

XX Sequence 1011 BP; 217 A; 252 C; 202 G; 340 T; 0 other;

#### alignment\_scores:

Quality:	Ratio:	Length:	Gaps:
918.00	3.600	301	0
Percent Similarity: 84.718	Percent Identity: 58.804		

alignment\_block:

US-09-755-017-2 x AAH32306 ..

Align seg 1/1 to: AAH32306 from: 1 to: 1011

```

5 AsnAspSerIleIleGlnLupheIleLeuLeuGlyPheSerAspArgpr 21
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 AATGCAAGTTTGAAGACCTTCTTATCTGATCTGATTTTCTACGCGC 59
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 oTrrLeuGluPheProLeuValAlaPheLeuIleSerTyrThrValT 38
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 TCATCTGGAAGTACTTCTCTGCTGATCTGATCTTCTGATGATTA 109
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 hTrrLeuGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLys 54
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 CACTGATGAGAAACCTGTCATCATCATCATCTGCAACCTGACCTCCAT 159
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 LeuHisThrProMetTyrPhePheLeuThrAsnLeuSerLeuAspLe 71
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 CTCACACCTCCCATGACTCTTCTTCTTCAAACTCTCATTTTGTGATC 209
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 uCysTyrThrThrCysThrValProGlnMetLeuValAsnLeuCysSer 88
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 CTCCTACACACACAGCTATCTCTCTGCTGAGTGGTGAATCTGCGGCC 259
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 LeArgLysValIleSerTyrArgGlyCysValAlaGlnLeuPheIlePhe 104
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 CGGAAAGACATCTCTTATGCTGCTGTAACATCACTTACTTCTTGT 309
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 LeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSerPheAs 121
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 CTCGCACTGGGAAACCCGACAGTGTCTCTGCTGCTGCTGCTGCTGCT 359
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 ParGpHeValAlaIleCysArgProLeuHisTyrSerValIleMetHis 138
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 TCGTATGACAGTGTGTGAGACCTTGCATTCACATGCTCTCATGACAC 409
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 InaArgLeuCysLeuGlnLeuAlaAlaIleSerTrpValThrGlyPheSer 154
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 CTCGTTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 459
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 AsnSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeuCysAsp 171
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 ACCTCAAGCACTTCTCTCTCTTCTTCTGATACCTCATGCTGATGACA 509
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 oTrrValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeus 188
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 TCGCCTAGTGAATCACTTCTTCTGTAAGCTCAGACCTCTCGATAT 559
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 eTrrCysValGluThrThrAlaAsnGluAlaGluLeuPheLeuValSerGlu 204
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 CAGGTGTGTAACCTTAGGCAATGAGCTGACCTCAAGTCAAGTACAGCTCC 609
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 LeuPheHisLeuLeuProLeuThrLeuIleIleIleSerTyrAlaPheIle 221
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 ATTCTTCTTCTCATCTCTCATCTCATCTCTCTCTCTCTCTCTCTCT 659
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 eValArgAlaValAlaLeuArgIleGlnSerAlaGluGlyArgGlnLysAla 238
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 TGGCCGGGCTGTACTGATGATGATCAACCACTGGGCTTCAAGAAAGTGC 709

```

238 heglthyrCysGlySerHisLeuIleValValSerLeuPheTyrSerThr 254  
 710 TTAGGACATGTGGAGCCCATCTATGATGTATCTCTTTTCATTCACA 759  
 255 AlValSerValTyrLeuGlnProPheSerPheSerTyrLysAspGlnI 271  
 760 GTATGTGCATGTATCTCCAGCCACCAACAAATTCCTCAAGATCAAG 809  
 271 yLysMetValSerLeuPheTyrGlyIleIleAlaPheMetLeuAsnProL 288  
 810 CAGTTCATTTGCCCTCTTTACACCTGTGTACACCTAGTCTTACCCCTC 859  
 288 euileTyrThrLeuArgAsnLysGluValLysGluGlyPheLysArgLeu 304  
 860 TAACTACACTTTCAGAAACAGAGATGTAGAGGGCAGTGAAGAGACTA 909  
 305 Val 305  
 910 ATG 912  
 seq\_name: /SID52/gcgsdata/geneseq/geneseqn/NA2001.DAT:AAH32314  
 seq\_documentation\_block:  
 ID AAH32314 standard; DNA; 936 BP.  
 AC AAH32314;  
 DT 30-JUL-2001 (first entry)  
 DE Human olfactory receptor polynucleotide, SEQ ID NO: 887.  
 XX  
 KW Human: olfactory receptor; OR: primary scent determination;  
 KM secondary scent determination; polypeptide library; odour receptor;  
 KN scent profile; scent fingerprint; scent representation; ds.  
 OS Homo sapiens.  
 PN MO200127158-A2.  
 PD 19-APR-2001.  
 PF 06-OCT-2000; 2000WO-US27582.  
 PR 08-OCT-1999; 99US-0158615.  
 PR 24-FEB-2000; 2000US-0184809.  
 XX  
 PA (DIGI-) DIGISCENTS.  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;  
 DR WPI: 2001-290713/30.  
 PT New polynucleotides which encode polypeptides involved in olfactory  
 PS sensation for identifying olfactory agonists and antagonists -  
 XX  
 XX Claim 8; Page 534; 1857pp; English.  
 CC The present sequence is one of a number of isolated polynucleotides  
 CC which encode polypeptides involved in olfactory sensation. The  
 CC polynucleotides can be used in screening for olfactory agonists and  
 CC antagonists. The methods allow for the determination of primary  
 CC scents and the identification of the odour receptors used to detect  
 CC these primary scents. The methods also enable determination of  
 CC secondary scents and the identification of combinations of odour  
 CC receptors that are involved in detecting such secondary scents.  
 CC This enables the construction of a scent representation (also called  
 CC a scent fingerprint or scent profile), which may be used to re-create  
 CC and edit scents. Libraries of olfactory receptors are useful for  
 CC determining the interaction pattern of a composition with the receptors,  
 CC and can be used for determining differences in the olfactory faculties  
 CC of different individuals.

XX  
 SQ Sequence 936 BP; 199 A; 236 C; 193 G; 308 T; 0 other;  
 alignment\_scores:  
 Quality: 912.00 Length: 294  
 Ratio: 3.707 Gaps: 0  
 Percent Similarity: 83.673 Percent Identity: 60.204  
 alignment\_block:  
 US-09-755-017-2 x AAH32314 ..  
 Align seq 1/1 to: AAH32314 from: 1 to: 936  
 12 PheIleuLeuGlyPheSerAspArgProIleuGlnIlePheProleuLe 28  
 37 TTTATTTCTACTTGGATTTTCTAATGTGGCTTCAGCTGAGAGTAGTCTCT 86  
 28 uValValPheLeuIleSerTyrThrValThrIlePheGlyAsnLeuThrI 45  
 87 TGTGTTATCTTGTATCTTACCTGATGACATGACATGACAGAAACCTGTTC 136  
 45 IeIleLeuValSerArgLeuAspThrLysLeuHisThrProMetTyrPhe 61  
 137 TCATCATCTCTGTATAGTGTGACTCCCATCTCCACACCAACATGTACTTC 186  
 62 PheLeuThrAsnLeuSerLeuAspLeuAspLeuCysTyrThrPheThrVa 78  
 187 TTCCTTTCAAACCTTCATTTCTGTGATCTGTGCACACACACAGCTCTAT 236  
 78 LProGlnMetLeuValAsnLeuCysSerIleArgLysValIleSerTyrA 95  
 237 CCTCAGTTCCTGCTGATGATCTCGGGGCGCGGAAACACATCTCGATG 286  
 95 rgGlyCysValAlaGlnLeuPheIlePheLeuAlaLeuGlyAlaThrGlu 111  
 287 CTGGTTCATGCTTCAACTTACTTGTCTGTCCACATGCGGAATCGCAG 336  
 112 TyrLeuLeuLeuAlaValMetSerPheAspArgPheValAlaIleCysAr 128  
 337 TGTGTCTACTGCTGATGATGCTGATGATGATGATGATGATGATGATG 386  
 128 gProLeuHisTyrSerValIleMetHisGlnArgLeuCysLeuGlnLeuA 145  
 387 ACCTTTCATTTACATGCTGCTCATGACACCTCGTTTCTGCACTGTGTG 436  
 145 IAlaIleAspTyrValThrGlyPheSerAsnSerValTyrLeuSerThr 161  
 437 CTGCGGCTCTGTGGTAAATGCTTTTACTATGTCACAGCACTTCATCTCC 486  
 162 LeuThrLeuGlnLeuProLeuCysAspProTyrValIleAspHisPheLe 178  
 487 TTTACTTTCGCGTACCCCTTGTGACATGCGCTGATGATGATGATGATG 536  
 178 uCysGluValProAlaLeuLeuLysLeuSerCysValGluThrThAlaAla 195  
 537 CTGTGAAGTTCACGACACTTCTGCGTTATCATGTGTGACACCCATGCCA 586  
 195 snGluAlaGluLeuPheLeuValSerGluLeuPheHisLeuIleProLeu 211  
 587 ATGACGTGACCCCTCATGATGATGATGATGATGATGATGATGATGATG 636  
 212 ThrLeuIleLeuLeuSerTyrAlaPheIleValAlaValAlaLeuArgI 228  
 637 ATTCTATTCCTACCTGCTATGCTGATGCTGCGGCTGTACTGAGCAT 686  
 228 eGlnSerAlaGluGlyArgGlnLysAlaPheGlyThrCysGlySerHisL 245  
 687 GCATATACACACAGGCGCTTCAGAAAGCTGTTAGACATGTGAGGCCATC 736  
 245 euileValValSerLeuPheTyrSerThrAlaValSerValTyrLeuGln 261  
 737 TTATGCTGTATCTCTTTTCATTCACAGTATGTCATGTGATCTCCAG 786



288 euileTyrrhrLeuArgAsnLysGluValLysGluGlyPheLysArgLeu 304  
|||||  
869 TGATCTATACACTAAGAAACAAAGATGTAAAGAGGCCCATGAAGAAGTGC 918  
305 ValAlaArg 307  
919 CTAGGGAAG 927



OM of: US-09-755-017-2 to: EST:\*. out\_format: pfs  
 Date: Feb 26, 2002 5:25 AM

About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL-frame+ .p2n.model -DEV-rlh  
 -Q-cg92.1/USPTO.spool/US09755017/runal\_25022002.165640.20760/app\_query.fasta\_1.374  
 -DB-EST -DBMT-fastap -SUFFIX-p2n.rst -GAPOP-12.000 -GAPEXT-4.000  
 -MIMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000 -OGAPOP-6.500  
 -OGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500 -FGAPOP-6.000  
 -FGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500 -DELDP-6.000  
 -DELEXT-7.000 -START-1 -MATRIX-blosum62 -TRANS-human40.cdi  
 -LIST-45 -DOCALLIGN-200 -THR\_SCORE-pct -THR\_MAX-100 -THR\_MIN-0  
 -ALIGN-15 -MODEL-LOCAL -OUTFMT-pfs -NORM-ext -HEAPSIZE-500  
 -MILEN-0 -MAXLEN-2000000000 -USER-US09755017.gcgn1.1.3653  
 -NCPU-6 -ICPU-3 -LONGLOGS -DEV\_TIMEOUT-120 -WARN\_TIMEOUT-30  
 -NO\_XLPHY -WAIT -THREADS-1

## Search information block:

Query: US-09-755-017-2

Query length: 313

Database: EST:\*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1205.400000

## score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_hlc:AK016560	+	605.50	933.98	2.7e-42	3063	AK016560 Mus musculus adult ma
gb_hlc:AK017036	+	604.50	930.01	1.2e-42	1394	AK017036 CIT-HSP-2354D1.TF CIT
gb_gss:AK0077154	-	573.00	891.18	1.8e-40	479	AK0077154 CIT-HSP-2354D1.TF CIT
gb_gss:AK0090606	-	567.50	879.04	8.5e-40	692	AK0090606 RPT-24-37B20.TV RPT
gb_gss:BH111304	+	566.50	876.23	1.2e-39	788	BH111304 RPT-24-36786.TV RPT
gb_gss:AK0068227	+	554.00	856.70	1.1e-38	642	AK0068227 2M024J24R Mouse 10kb
gb_gss:AK0255734	+	554.00	841.91	9.9e-38	726	AK0255734 RPT-23-16619.TV RPT
gb_hlc:AK016338	+	544.00	834.90	2.4e-37	1501	AK016338 Mus musculus adult me
gb_est1:AI604336	-	534.00	829.48	1.1e-34	794	AI604336 vv74c09.xl Stratagene
gb_gss:AK02913406	-	509.50	787.22	1.1e-34	764	AK02913406 RPT-23-165C12.TV RPT
gb_gss:AK0253949	-	505.50	782.53	2.0e-34	797	AK0253949 IM009J01F Mouse 10kb
gb_gss:AK02909618	-	497.50	766.46	1.2e-33	797	AK02909618 RPT-24-222G18.TV RPT
gb_est2:BG193339	+	493.00	761.87	3.8e-33	762	AK0193339 RST12467 Athersys RAGE
gb_gss:AK0260393	+	492.00	760.55	4.9e-33	740	AK0260393 IM0429N1SR Mouse 10kb
gb_est2:BG197640	+	479.00	743.46	3.0e-32	534	AK0197640 RST17016 Athersys RAGE
gb_gss:AK02664867	+	475.00	735.59	8.3e-32	632	AK02664867 IM054E05R Mouse 10kb
gb_gss:AK02511623	-	470.50	727.78	2.3e-31	686	AK02511623 IM035C17R Mouse 10kb
gb_gss:AK02686625	+	468.50	724.53	3.4e-31	695	AK02686625 RPT-23-26E4.TV RPT
gb_gss:AK0274496	+	467.00	724.49	3.4e-31	548	AK0274496 2M024JB07F Mouse 10kb
gb_gss:AK0271125	+	466.50	723.62	3.8e-31	553	AK0271125 RPT-23-146J4.TV RPT
gb_gss:AK0482125	+	465.00	719.78	6.3e-31	646	AK0482125 RPT-11-261E3.TV RPT
gb_gss:AK02638594	+	463.00	716.19	1.0e-30	678	AK02638594 IM049E06R Mouse 10kb
gb_gss:AK02964406	+	460.00	711.45	1.8e-30	682	AK02964406 2M023A003R Mouse 10kb
gb_gss:AK02111477	-	457.00	705.95	3.7e-30	743	AK02111477 RPT-23-9A5.TV RPT
gb_gss:AK0078130	-	456.00	705.59	3.9e-30	656	AK0078130 RPT-24-164M20.TV RPT
gb_gss:AK0103967	-	455.00	704.71	4.3e-30	611	AK0103967 RPT-23-33G18.TV RPT
gb_gss:AK0270443	-	450.00	698.43	9.7e-30	622	AK0270443 RPT-24-161N23.TV RPT
gb_est1:AV688362	-	448.50	693.55	1.8e-29	619	AV688362 AV688362 GKC Homo sapi
gb_est1:AI1448824	+	448.00	693.67	1.8e-29	521	AI1448824 GCT08a02.1J Scars plac
gb_gss:AK0407823	+	446.00	692.21	2.2e-29	521	AK0407823 IM017B09R Mouse 10kb
gb_gss:AK0252656	+	446.00	688.67	3.4e-29	732	AK0252656 RPT-23-210N3.TV RPT
gb_gss:AK0191257	-	445.50	684.61	3.0e-29	629	AK0191257 RPT-23-266H21.TV RPT
gb_est1:AV693686	-	442.00	683.52	5.0e-29	534	AV693686 AV693686 GKC Homo sapi
gb_gss:AK012830	+	442.00	683.73	5.0e-29	534	AK012830 IM002J19R Mouse 10kb
gb_gss:AK02697150	-	436.00	675.10	1.9e-28	609	AK02697150 RPT-23-231B2.TV RPT

gb\_gss:BH069789 - 436.00 673.78 2.3e-28 698 | BH069789 RPT-24-331A10.TVB  
 gb\_gss:AK0648626 + 435.00 672.18 2.5e-28 632 | AK0648626 IM0517B18R Mouse 1  
 gb\_gss:AK0725747 - 434.50 669.63 3.9e-28 842 | AK0725747 RPT-24-98112.TVB  
 gb\_gss:BH078443 - 433.50 670.73 3.4e-28 639 | BH078443 RPT-24-34969.TV R  
 gb\_est1:AV689607 - 433.50 670.13 3.7e-28 680 | AV689607 AV689607 GKC Homo

seq\_name: gb\_hlc:AK016560

seq\_documentation\_block:

LOCUS AK016560 3063 bp mRNA HTC 05-JUN-2001  
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932441H21, full insert sequence.

ACCESSION AK016560

VERSION AK016560.1 GI:12855357

KEYWORDS CAP trapper.

Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library clone:4932441H21.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 3063)

Carninci,P., and Hayashizaki,Y.

High-efficiency full-length cDNA cloning

Methods in enzymology. 303, 19-44 (1999)

99279253

10349636

2 (bases 1 to 3063)

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

Genome research. 10 (10), 1617-1630 (2000)

20499374

11042159

3 (bases 1 to 3063)

Shibata,K., Itoh,M., Aizawa,K., Nagoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,Y., Nishi,K., Katsuna,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsuno,H., Sakaguchi,S., Ikegami,T., Kasaiyagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneeda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multiplexed sequencer

Genome research. 10 (11), 1757-1771 (2000)

11076861

4 (bases 1 to 3063)

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 3063)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

Submitted (10-Jun-2000) Yoshida Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)





Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [mdadams@tigr.org](mailto:mdadams@tigr.org)  
Clones are available from Research Genetics ([inforesgen.com](http://inforesgen.com)). BAC  
end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: M3-21

seq_name:	gb_gss:AZ090606	,	
seq_documentation_block:			
LOCUS	AZ090606	692 bp	DNA
DEFINITION	RPcI-23-27B20_TV RPcI-23 Mus musculus genomic clone RPcI-23-27B20		
	DNA sequence.		
ACCESSION	AZ090606		
VERSION	AZ090606.1	GI:7732649	
KEYWORDS	GSS.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		



/clone\_11b="RPCI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTARAC1: Site.1: BamHI: Site.2: BamHI:  
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  
 library was cloned in the pTARAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."  
 BASE COUNT 168 a 217 c 162 g 241 t  
 ORIGIN

alignment\_scores:  
 Quality: 566.50 Length: 265  
 Ratio: 2.737 Gaps: 7  
 Percent Similarity: 78.113 Percent Identity: 46.415

alignment\_block:  
 US-09-755-017-2 x BH11304 ..

Align seg 1/1 to: BH11304 from: 1 to: 788

```

35  TTTTThValThrllePhelglnLeuThrlleLeuValSerArgle 51
    ||| : : : : : ||| : : : : : |||
6  TACATCTTCAGCCTCTCGCAATGGAATAGTAGTATCTTCCT 55
    ||| : : : : : ||| : : : : : |||
51  uasprThlLeuHsthrProketyrPhelLeuThrlleLeuSerl 68
    ||| : : : : : ||| : : : : : |||
56  GACACCCAGCTGCACACCCCTATGACTCTTCTTCTTCATCTG 105
    ||| : : : : : ||| : : : : : |||
106 TCTTGACATATCTATGCTTCACACATGTCGCCAAGATGTTGCC 155
    ||| : : : : : ||| : : : : : |||
85  LeuGysSerlleArglyVallleSerTyArglyCysValAla 101
    ||| : : : : : ||| : : : : : |||
156 CTTATTAACAGAGAAAGAACATCTATGCTGCCCATGCATCAC 205
    ||| : : : : : ||| : : : : : |||
101 upheillePhelLeuAlaLeuGlyValThrlleLeuValAla 118
    ||| : : : : : ||| : : : : : |||
206 ATTCTTGATTTGGCTTGTGGCCCTTGGAGTCTTGATTTGGCA 255
    ||| : : : : : ||| : : : : : |||
118 eSerPhespargPheValAlaileCysArgProleuHsthrSer 134
    ||| : : : : : ||| : : : : : |||
256 TGCACATGACAGAGTTGTGGCCATCTGCACCCCTACATCACT 305
    ||| : : : : : ||| : : : : : |||
134 lleMetHsGlnArgLeuGysLeuGlnLeuAlaAlaSerTrp 151
    ||| : : : : : ||| : : : : : |||
306 CATCATGAGCTTGGAAGATGTGTGCTGCCCTGCTGCATCT 355
    ||| : : : : : ||| : : : : : |||
151 hGlyPheSerAsnSerValTrpLeuSerThrlleLeuGlnLeu 167
    ||| : : : : : ||| : : : : : |||
356 GTGGATTTAGCCTCTCTGTGGACACACATTTCTGCTAAGTT 405
    ||| : : : : : ||| : : : : : |||
168 LeuGysAspProTyVallleAspHsPheLeuGysGluValPro 184
    ||| : : : : : ||| : : : : : |||
406 TTTTGGGGCCCGAGAAATATAACACCTCTGTGAATCTGCTGT 455
    ||| : : : : : ||| : : : : : |||
184 uLeuLysLeuSerCysValGluThrlleAsnGlnAlaGluLeu 201
    ||| : : : : : ||| : : : : : |||
456 TCTCAAGCTGGCCCTGCGCACACTTGTATCAACCAATC... 502
    ||| : : : : : ||| : : : : : |||
201 euValSerlLeuPheHsLeulle...ProleuThrlleu... 216
    ||| : : : : : ||| : : : : : |||
503 TACGTGCAGTGTCTTGTAGTTGAGACCCCTTGTCACTGTGT 552
    ||| : : : : : ||| : : : : : |||
216 eSerTyAlaPheHsLeuValArgAlaValLeuArglleGlnSer 233
    ||| : : : : : ||| : : : : : |||
553 ATCTTACAGCACATCTCTGACATCTCTGAATAATGCTCAAGAG 602
    ||| : : : : : ||| : : : : : |||
233 lYArgGlnLysAlaPheGlyThrCysGlySerHsLeuValSer 249
    ||| : : : : : ||| : : : : : |||
603 GTGCGAGGAAGGCTTTTCACCTGCTCCCATCTGTGTGGTGG 652
  
```

```

250  LeuPheTySerThrlleValSerValTyTrleuGlnProSerPro 265
    ||| : : : : : ||| : : : : : |||
653 CTCTTCTTGTATACCATGCTGTGTTATATAGTCCCTGACTGT 702
    ||| : : : : : ||| : : : : : |||
266 SerSerLysAspGlnGlyMetValSerLeuPheTyTrlleAla 282
    ||| : : : : : ||| : : : : : |||
703. CACGAGAGAGAGAGAGAAATTTGTCTCAGTCCACAGTCTTTT 752
    ||| : : : : : ||| : : : : : |||
282  aprometLeuAsnProleuTyTrlleuArgAsn 294
    ||| : : : : : ||| : : : : : |||
753 CCAAGTGAATTCCT...CATATAGTCTGAGGAGAC 786
    ||| : : : : : ||| : : : : : |||
seq_name: gb_gss:A2969227
  
```

seq\_documentation\_block: 642 bp DNA 27-APR-2001  
 LOCUS A2969227  
 DEFINITION 2M0241J24R Mouse 10kb plasmid U0GC2M 1library Mus musculus genomic  
 clone U0GC2M0241J24 R, DNA sequence.  
 ACCESSION A2969227  
 VERSION A2969227.1 GI:13840454  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 642)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellily  
 M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah  
 Department of Biology  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0241 row: J column: 24  
 Seq primer: CACACAGGAACAGCTATGACAC  
 Class: plasmid ends  
 High quality sequence stop: 642.  
 Location/Qualifiers  
 1. 642  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="U0GC2M0241J24"  
 /clone\_11b="Mouse 10kb plasmid U0GC2M 1library"  
 /sex="Female"  
 /lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PMD42uv; Purified genomic DNA from M.  
 musculus C57BL/6J (female) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PMD42 (9114732114[9b]AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into

#### FEATURES

source

chemically-competent *E. coli* XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 130 a 166 c 129 g 217 t  
ORIGIN

alignment\_scores: Length: 213  
Quality: 554.00 Gaps: 0  
Ratio: 3.240  
Percent Similarity: 80.282 Percent Identity: 47.418

alignment\_block:  
US-09-755-017-2 x AZ969227 ..

Align seg 1/1 to: AZ969227 from: 1 to: 642

```

79 ProGlnMetLeuValAsnLeuCysSerIleArgLysValIleSerTyrAr 95
   |||||.....|.....|.....|.....|.....|.....|.....|
   3 CCCAGATGATGTCATCTTCTGTCAGAGAAATCATTTCTCTATGG 52
95 gglCysValIleGlnLeuPheIlePheValAlaLeuGlyAlaThrGluT 112
   |||||.....|.....|.....|.....|.....|.....|.....|
53 AGGCTGTGTGACCAAGCTCTTTCATTTCTTTGTGGCTCAGAGT 102
112 yrlLeuLeuAlaValMetSerPheAspArgPheValAlaIleCysArg 128
   |||||.....|.....|.....|.....|.....|.....|.....|
103 GTCCTCTCTGGCAGCATGTCATATGATATATTCCTATCTGTAG 152
129 ProLeuHisTyrSerValIleMetHisGlnArgLysLeuGlnLeuAl 145
   |||||.....|.....|.....|.....|.....|.....|.....|
153 CCGTTAAGGATCATTTATATGACAAAGCTCTGTCAGCTGCTAGC 202
145 aAlaAlaSerTyrValThrGlyPheSerAsnSerValTrpLeuSerThr 162
   |||||.....|.....|.....|.....|.....|.....|.....|
203 AGCTTCATGCTGACAGCTGGGTTTCTCACTGATGTCACACAGGTTT 252
162 eurThrLeuGlnLeuProLeuCysAspProTyrValIleAspHisPheLeu 178
   |||||.....|.....|.....|.....|.....|.....|.....|
253 TGACCTTCACACTGGCCCTTTGTGTGTACAAATCATATTTCTTC 302
179 CysGluValProAlaLeuLeuLysLeuSerCysValGluThrThrAla 195
   |||||.....|.....|.....|.....|.....|.....|.....|
303 TGTGCAATACCTCCCTTCATCTCTGTCGATACATCTCCCTCAA 352
195 nglAlaLeuLeuPheLeuValSerGluLeuPheHisLeuLeuProLeu 212
   |||||.....|.....|.....|.....|.....|.....|.....|
353 TGAATGCTTTGCTGTCATTTGGGATCCATAGGCTGACCTCTTC 402
212 hrLeuIleLeuIleSerTyrAlaPheIleValArgAlaValLeuArg 228
   |||||.....|.....|.....|.....|.....|.....|.....|
403 TGTGGTATCTCTTCTCTATATGAGGAGTATCCACCATCTCGAGATC 452
229 GlnSerAlaGluGlyArgGlnLysAlaPheGlyThrCysGlySerHis 245
   |||||.....|.....|.....|.....|.....|.....|.....|
453 CGTTCCTCTGAGGGGAGGACAAAGCTTTCCACCTGCTGCCACCT 502
245 uIleValValSerLeuPheTyrSerThrAlaValSerValTyrLeuGln 262
   |||||.....|.....|.....|.....|.....|.....|.....|
503 GCTCATTTGTATTTCTCTATATGAGGAGTATCCACGATATGAGGC 552
262 rOpSerProSerSerLysAspGlnGlyLysMetValSerLeuPheTyr 278
   |||||.....|.....|.....|.....|.....|.....|.....|
553 CACATCTCATCTTCTCTAGAGAAAGATAGATTGATCTCACTGCTGAT 602
279 GlyIleIleAlaProMetLeuAsnProLeuIleTyrThr 291
   |||||.....|.....|.....|.....|.....|.....|.....|
603 AGTGTTCACACCCATGCTGAATCTGTAATTATATAGC 641

```

seq\_name: gb\_AZ255734

seq\_documentation\_block: 726 bp DNA GSS 26-JUL-2000  
LOCUS AZ255734  
DEFINITION RPCI-23-16619..TJ RPCI-23 Mus musculus genomic clone RPCI-23-16619,

ACCESSION DNA sequence.  
A2255734  
VERSION GI:9458784  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 726)  
AUTHORS Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akimov, B., Levins, M., McGann, S., Tesgaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.  
TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
COMMENT Other GSSs: RPCI-23-16619.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@eijong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 166 row: I column: 9  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..726  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-16619"  
/clone\_id="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1: EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI MethyIase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 167 a 171 c 118 g 270 t  
ORIGIN

alignment\_scores: Length: 239  
Quality: 544.00 Gaps: 0  
Ratio: 2.973  
Percent Similarity: 76.569 Percent Identity: 42.259

alignment\_block:  
US-09-755-017-2 x AZ255734 ..

Align seg 1/1 to: AZ255734 from: 1 to: 726

```

26 ProLeuLeuValAlaPheLeuIleSerTyrThrValThrIlePheGlys 42
   |||||.....|.....|.....|.....|.....|.....|.....|
5 CCTGTGTCTTCTGTCCTGTGATATGTAATAGTAAGTACGCGGAA 54
42 nLeuThrIleIleLeuValSerArgLeuAspThrLysLeuHisThrPro 59
   |||||.....|.....|.....|.....|.....|.....|.....|
55 TTGGGTTTGGAATCTGAATGGACTGAATGCTACCTTCACACCGGA 104
59 eTyrTrpPheLeuThrAsnLeuSerLeuLeuAspLysCysTyrThr 75
   |||||.....|.....|.....|.....|.....|.....|.....|
105 TGTACTTTTCTCTTAACTGTCTGTGATAGACCTGTGTACTCTTCA 154

```







## alignment\_scores:

Quality: 544.00 Length: 296  
 Ratio: 2.693 Gaps: 2  
 Percent similarity: 68.243 Percent identity: 42.905

alignment\_block:  
 US-09-755-017-2 x AK016338 ..

Align seg 1/1 to: AK016338 from: 1 to: 1501

```

5 AsnaSpSerIleIleGlnIupheIleLeuGlyPheSerAsparPr 21
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
615 AATATAACCAATAGAGAGATTCCTCCATGGGATTTCTGCACACCA 664
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
21 cTTPLeuGlnPhePLeuLeuValaIphLeuIleSerTyrThrValT 38
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
665 TGAGCTGCAGATCTACAGGGCTGCTCTCTTGTGACATACCTAGTCG 714
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
38 hTLePheGlyAsnLeuThrIleIleLeuValSerArgLeuSprThrLys 54
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
715 GCTCAGCAGGAGACATCATATTATACCATCAACACTAGACCCACAG 764
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
55 LeuHISThrProMetTyrPhePheLeuThrAsnLeuSerLeuLeuAspLe 71
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
765 CTCACAGTCTCCATATGATTACTTCTGAAGCA.CTTTCATTTCTGACCT 813
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
71 uCysTyrThrThrCysThrValProGlnMeLeuValaAsnLeuCysSerI 88
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
814 CTCATCCCTCTCTCTGTCACAGTCCACAGTATGTTGACATGTCCTCCG 863
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
88 LeArgLysValIleSerTyrArgGlyCysValaIleGlnLeuPheIlePhe 104
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
864 AAAGTGGCTACATTTTCATATGACACAGTGCATCTGACAGATTTTCTTC 913
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
105 LeuAlaLeuGlyAlaThrGlnTyrLeuLeuLeuAlaValMetSerPheAs 121
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
914 ACAGCTTTTCCCTGGGCTGAGCTGGCCATCTCACAGCATGCTTATATGA 963
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
121 PArgPheValAlaIleCysArgProLeuHISThrSerValIleMetHisG 138
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
964 CCGCATATGAGCCATCTGCTCCATTCGACATATGAGTCATCATGATGATC 1013
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
138 InArgLeuCysLeuGlnLeuAlaIleAlaSerTyrValThrGlyPheSer 154
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1014 CCAGAAAGTGACCTGGGCTGGCAACGTGTGCTAGTGAAGGTATAC 1063
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
155 AsnSerValTyrPLeuSerThrLeuThrLeuGlnLeuProLeuCysAspPr 171
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1064 TCAGGACACTTGTACATACAGTACACTCTTATTCACAGATTCTGTGGGA 1113
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
171 cTTPValIleAspPheIleLeuCysGlnValaProAlaLeuLeuLysLeu 188
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1114 CAAATATTATTCACAGATTCTGTGTGATGTCCCCAGTGTGCACAGCTCT 1163
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
188 eTcCysVal.....GlnThrThrAlaAsnGlnAlaGlnLeuPheLeuVal 202
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1164 CCGTCTTAATAGCACCTTGAATTAATGACATGGTTAGTTCCCGACT 1213
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
203 SerGlnLeuPheHISLeuIleProLeuThrLeuIleLeuIleSerTyrAl 219
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1214 GCAGTTTCTTGGCTGT.....TTCACTGGGATTGTCACTCCATATGT 1257
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
219 aPheIleValaArgAlaValaLeuArgIleGlnSerAlaGlnGlyArgGlnL 236
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1258 CCACATATTCTCCACAGTCTCAGAGATGCCCTGCTGTAAGAGAGGTCTA 1307
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
236 yAlaIlePheGlyThrCysGlySerHISLeuIleValaSerLeuPheTyr 252
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1308 AGGTTCTTCTTACTGCTGCCCACTTTTGTGTGTCATTTGTTCTT 1357
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
253 SerThrAlaValSerValTyrLeuGlnProProSerProSerSerLys 269

```

```

1358 TCTACAGGGCCCTTTCATATCTAAACCTAACCTACCTCTCACTCG 1407
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
269 pGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaProMetLeuA 286
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1408 TTTAGAGTTCTTGTCTCTATCTTTACACAGTACTACCTCCAACTCTCA 1457
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
286 snProLeuIleTyrThrLeuArgAsnLysGlnValLys 298
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1458 ACCCTGTATTATTAGTCTGAGAAATGACGATTAAG 1495
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

seq_name: gb_est1:AI604386

seq documentation_block:
LOCUS AI604386 523 bp mRNA EST 21-APR-1999
DEFINITION v74c09.x1 Striatagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1228144 3' similar to SW:01F2_HUMAN Q15062 OLFACTORY
RECEPTOR- LIKE PROTEIN FAT11. ;, mRNA sequence.
ACCESSION AI604386
VERSION AI604386.1 GI:4613553
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theisling,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Rittler
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
TITLE Unpublished (1999)
JOURNAL Contract: Marra M/WashU-NCI Mouse EST Project 1999
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:653736
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone, similarity on wrong strand
High quality sequence stop: 457.
location/Qualifiers
1..523
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1228144"
/clone_lib="Striatagene mouse skin (#937313)"
/sex="females"
/tissue.types="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcORI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb. Uni-ZAP XR Vector: -5'
adaptor sequence: 5' GAAATCGGCACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTTCTTTT 3'"
BASE COUNT 149 a 133 c 150 g 89 t 2 others
ORIGIN
1
alignment_scores:
Quality: 534.00 Length: 172
Ratio: 3.608 Gaps: 0
Percent Similarity: 86.047 Percent Identity: 58.140
alignment_block:

```



```

176 sPheLeuGysGlValProAlaLeuLeuLysLeuSerCysValGluThrT 193
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
312 TTCTACTGATGATGTCCTCCAGCATCATCAACTTGGCTGCACAGACATCT 263
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
193 hAlaAsnGluAlaGluLeuPheLeuValSerGlu.....LeuPheHis 207
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
262 TTGTT.....CTTGAGCTGCTGATGATGATTTCCAAATGATGCTGTTGCT 219
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 LeuLeuProLeuThrLeuLeuLeuLeuSerTyrlaPheLeuValArgAl 224
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
218 ACTCTGCTGTTGCTGTTGCTGCTGCTCATACACA...GTCAATCTTAT 172
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
224 aValLeuArgIleGlnSerAlaGluGlyArgGlnLysAlaPheGlyThrC 241
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 GATGCTTGAATCTCATTTGAGAGAGGCGAGGAAGAAGCCATCTCCACT 122
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ySGLSerHisLeuLeuValValSerLeuPheTyrlSerThrAlaValSer 257
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 GCACCTCCCACTACATGCTGTCACACGCTGCTGCTGCTGCTGCTAT 72
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
258 ValTyrlLeuGlnProProSerProSerSerLysAsnGlyLysMetAla 274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 GTCTATGCCAGGCGCCCTTCACTGCCCTCCCGCCAGGAT.....AAGCCAT 28
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
274 lSerLeuPheTyrlGlyIleIleAla 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
27 CTCTGTCACTTCACAGTCATCTCC 3
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: gb_gss:A2353949

seq_documentation_block:
LOCUS      A2353949             676 bp      DNA             GSS             02-OCT-2000
DEFINITION IM0093J01F Mouse 10kb plasmid UUCG1M library Mus musculus genomic
ACCESSION  A2353949
VERSION    A2353949.1 GI:10464961
KEYWORDS   GSS.
SOURCE      house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 676)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
JOURNAL    Contact: Robert B. Weiss
COMMENT    University of Utah
            University of Utah Genome Center
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0093 row: J column: 01
            Seq primer: CGTGTAAACGACGCGCCAGT
            Class: plasmid ends
            High quality sequence stop: 676.
            Location/Qualifiers
                1..676
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUCG1M0093J01"
                /clone_1kb="mouse 10kb plasmid UUCG1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: pMD42uv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson

```

```

Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (9114/732114/9b/AF12902.1), a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT      173 a      149 c      135 g      219 t
ORIGIN

alignment_scores:
    Quality: 505.50      Length: 232
    Ratio: 2.939      Gaps: 3
    Percent Similarity: 74.138      Percent Identity: 46.552

alignment_block:
US-09-755-017-2 x A2353949 ..

Align seg 1/1 to: A2353949 from: 1 to: 676

74 ThrThrCysThrValProGlnMetLeuValAsnLeuCysSerIleArgly 90
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 TCCTCACTACACAGCACCCAAATGATGTTGACCTGTTACTGAGAAGAAA 52
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
90 sValIleSerTyrlArgGlyCysValAlaGlnLeuPheIlePheLeuAla 107
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 GACTATATCTCATGCTGGGGTGCATGTA.CAACCTTTGGGGTCATTTCT 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
107 euGlyAlaThrGlyTyrlLeuLeuLeuAlaValMetSerPheAspArgPhe 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 TTGCTGACATGAGATCTTCACTTACTGCTGATGCGCTATGATAGTAT 151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 ValAlaIleCysArpProLeuHisTyrlSerValIleMetHisGlnArgle 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 GTGGCCATTTGTAACCTCTCCACTATATGACATATGAGCGGAGAG 201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
140 uCysLeuGlnLeuAlaAlaAsnTrpValThrGlyPheSerAsnSerV 157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 ATGCATATAGATGTTGCTGCGAAGCATGATGATGAGCTGTTTACATTTCTA 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
157 alrPleuSerThrLeuThrLeuGlnLeuProLeuCysAspProTyrlVal 173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 TTAATCCAGAGGCTTTGCTGTCACCTCCCATTTTGTGGACCAATAGAG 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 lIleAspHisPheLeuGysGluValProAlaLeuLeuLysLeuSerCysVa 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 ATTGATGACATATTTCTGTGATGATCATCTGTAAGTGAACCTGACGAC 351
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
190 lGluThr.....ThrAlaAsnGluAlaGluLeuP 200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 GCACACTTACATTTGCTGCTGTTGTTGACAGCGCMACAGGCGACCATTTG 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 heLeuValSerGluLeuPheHisLeuLeuProLeuThrLeuIleLeuIle 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 CATTTGGGAAGTTTTCATCTGCTGATGTCATACACAGCATTTTCATATA 451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 SerTyrlAlaPheIleValArgAlaValLeuArgIleGlnSerAlaGlu 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 TCT.....CTGAGAAGCAGTCATCCGAGG 477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 yArgGlnLysAlaPheGlyThrCysGlySerHisLeuIleValValSerL 250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

478 CAGACGCAAGCTCTCTCCACTTGTGATCTCACATTGCTGTGTATCA 527  
 250 euphryserThralaValSerValThreulinProProserProser 266  
 528 TTTTGTGGCCCTCTACTTTATGATATGCGGCTCAGACTACCTTC 577  
 267 SerLysAspGlnGlyMetValSerLeuPheTyrglyIleIleAlaP 283  
 578 TCTGAGAC.....AAGATGAGAGTATATTACACTATTTACTCTC 621  
 283 OmethleuAsnProleulinIleTyThreuarGanLysGluValLys 298  
 622 CAGCTGAATCCCTCTAATTACCTCTAGAAATGACAGATAAG 667

seq\_name: gb\_gss:A2909618

seq\_documentation\_block: 797 bp DNA GSS 05-MAR-2001  
 LOCUS A2909618 RPI-24-222G18.rj RPI-24 Mus musculus genomic clone RPI-24-222G18  
 DEFINITION RPI-24-222G18.rj RPI-24 Mus musculus genomic clone RPI-24-222G18  
 , DNA sequence.  
 ACCESSION A2909618  
 VERSION A2909618.1 GI:13228563  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 797)  
 Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akhurst, B., Levins, M.,  
 Tsagaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, F.,  
 Russell, D., de Jong, P. and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPI-24  
 Unpublished (1999)  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (http://www.choi.choi.org/bacpac/orderingframe.htm). BAC end  
 page: http://www.tigr.org/tigr.org/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 222 row: G column: 18  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers  
 source  
 1..797  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPI-24-222G18"  
 /clone\_1lb="RPI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTARBAC1. Site 1: BamHI. Site 2: BamHI. The  
 RPI-24 Mouse BAC library produced by Pieter de Jong. The  
 library was cloned in the pTARBAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."  
 BASE COUNT 225 a 167 c 243 g 162 t  
 ORIGIN

alignment\_scores:  
 Quality: 497.50 Length: 266  
 Ratio: 2.578 Gaps: 5  
 Percent Similarity: 72.556 Percent Identity: 41.729

alignment\_block:  
 US-09-755-017-2 x A2909618/rev

Align seg 1/1 to reverse of: A2909618 from: 1 to: 797

39 IlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLys 55  
 795 CTGCTAGGCAACCTGCTCATGATGACACCATGATGAGTCAACACACTCT 746  
 55 uhlsthr.PrometTyrrPhePheLeuThrAsnLeuSerLeuAspLeu 71  
 745 CCACACACCCCATGATGATGATGATGATGATGATGATGATGATGATG 636  
 72 CysTyrrThrCysThrValProGlnMetLeuValAsnLeuCysSer 88  
 695 TTCTACACCTGTGTCATCATCCACCATGATGATGATGATGATGATGAT 646  
 88 eArgLysValIleSerTyrrArgGlyCysValAlaGlnPheIlePhe. 104  
 645 ACCTTCATCCATGCGCTTCTGCGCTGCGCTGCGCTGCGCTGCGCTCT 596  
 105 LeuAlaLeuGlnValThrGluTyrrLeuLeuValAlaMetSerPheAs 121  
 595 TTTCATATTGGCTTTCACCCATTCCTTTCTACTACCTCATGGGTTATGA 546  
 121 parpheValAlaIleCysArgProLeuThrIleSerValIleMetHis 138  
 545 CCGCTAGTGGCCATGCTGTCACCCACATGATGATGATGATGATGATGAT 496  
 138 InArgLeuCysLeuGlnLeuAlaAlaSerThrValThrGlyPheSer 154  
 495 CCGCTAGTGGCCATGCTGTCACCCACATGATGATGATGATGATGATGAT 446  
 155 AsnSerValThrPheSerThrLeuThrLeuGlnLeuProLeuGlyAsp 171  
 445 ATGGGACATGCTGTCACCCACATGATGATGATGATGATGATGATGAT 396  
 171 oTyrrValIleAspHisPheLeuCysGluValProAlaLeuLysLeu 188  
 395 CAATGAGATCCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 346  
 188 eCysValGlnGluThrThrAlaAsnGluAlaGlu.....LeuPheLeu 201  
 345 CAGTCGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATG 296  
 202 ValSerGlnLeuPheHisLeuIleProLeuThrLeuIleLeuIleSer 218  
 295 ATCAGACGCCCTC.....CTGGGCTGCTTCTCTCTGATCTCTCTCAT 252  
 218 yralaPheIleValArgAlaValLeuArgIleGlnSerAlaGluGlyArg 234  
 251 ATGCTTCATGTTGTTTACCATCTTGAAGATACCATCAGCTGAGGGTCG 202  
 235 GlnLysAlaPheGlyThrCysGlySerHisLeuIleValValSerLeu 251  
 201 CACAGAGCTTCTCCACATGATGATGATGATGATGATGATGATGATGAT 152  
 251 eTyrrSerThrAlaValSerValTyrrLeuGlnProProSerProSerSer 268  
 151 TTATGCTTGTCTTGTGATTTACCTCAAGCTTAAGGGCCCAAGTCTC 102  
 268 yAspGlnGlyMetValSerLeuPheTyrglyIleIleAlaPromet 284  
 101 TGGAGAGAGATACTGATGATGATGATGATGATGATGATGATGATGAT 52  
 285 LeuAsnProleulinIleTyrrThrLeuArgAsnLysGluValLys 298  
 51 CTTAGTCCATCATCTTCACTGATGATGATGATGATGATGATGATGATG 10

seq\_name: gb\_est2:BG193339

seq\_documentation\_block: 762 bp mRNA EST 21-APR-2001  
 LOCUS BG193339 RST12467 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 DEFINITION RST12467 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG193339

seq\_name: gb\_gss:AZ607393

337 GACATGAGACATTTCTCTCTGGCTGTATGGCTTATGACCGCTTTGTGCG 386  
125 ailecysargproleuinhistyrservalilemethisglnarglyleucyl 142  
387 CGTGGCCGCCCTTACATTACACAGCAAGATGACCATCAGCTGTGTG 436  
142 euglnleualaialasertrpvalilthrglypheseranservaltrp 158  
437 CCGTGTGGTGTCTACGTGATCATGGGTGGTCCCACTGCAGTGTCTGTG 486  
159 leusertrhleuthrleuglnleuproleucysasprotyrvalilalas 175  
487 CACACCCCTGCTGATGCTGCTGACTTCATTCCTGTGCAGCAACACCATCCC 536  
175 phisphleucysgluvalproalaleuileuylsleusercysvalglut 192  
537 CCACATCTTCTGCGATGTGACGCCCTCTCGAAGCTCTCTGTGACACA 586  
192 hrthralasngluagluleupheleuvalsergluleuphehisleu 208  
587 CACACCTCAGTGAAGTGAATGATTCATTACGAGCGCCCTCATGTCAGATG 636  
209 ileproleuthrleuileuileusertryralaphellevalaralava 225  
637 ACCCATTTCTTTGCTCTGCTTCCTCATATGACATCACCCTGCGTGT 686  
225 lleuargileginseralaglulgiatrglnlalsalapheglythrlys 241  
687 CCTGAGGGCCAT...CCACAAAGGAGATGATAAGCCCTTCACACTGT 733  
242 glyserhisleuilevalvalserleu 250  
734 GSGTCCACACTGGCGAGGTCTCTCTC 760

seq\_name: gb\_gss:AZ607393

seq\_documentation\_block:  
LOCUS AZ607393 740 bp DNA GSS 13-DEC-2000  
DEFINITION IM0429N1SR Mouse 10kb plasmid UNGCM library Mus musculus genomic  
clone UNGCM0429N1S R, DNA sequence.  
ACCESSION AZ607393  
VERSION AZ607393.1 GI:11729583  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 740)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0429 Row: N Column: 15  
Seq primer: CACACAGAAACACCTATGACC  
Class: plasmid ends  
High quality sequence stop: 740.  
Location/Qualifiers  
1..740  
/organism="Mus musculus"  
/strain="C57BL/6J"

FEATURES  
SOURCE

/db\_xref="taxon:10090"  
 /clone="U08C1M0429N15"  
 /clone\_lib="Mouse 10kb plasmid U08C1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42ntv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 276 a 134 c 159 g 171 t  
 ORIGIN

alignment\_scores:  
 Quality: 492.00 Length: 244  
 Ratio: 2.718 Gaps: 0  
 Percent Similarity: 74.180 Percent Identity: 42.623

alignment\_block:

US-09-755-017-2 x AZ607393/rev ..

Align seg 1/1 to reverse of: AZ607393 from: 1 to: 740

```

30 ValPheLeuIleSerThrValThrIlePheGlyAsnLeuThrIleI 46
   ::::::::::: || ::::::::::: ||::| |::| |
733 AGTTTCTAGTGTGTTATCTTACTATATCGTCACAAATCTTGATGAT 644
   ::::::::::: || ::::::::::: ||::| |::| |
46 eLeuValSerAlaGluAspThrIleuHsrThrPrometTyrPhePhe 63
   ::::::::::: || ::::::::::: ||::| |::| |
683 CATTTTGATGACAAATGATGATCAGCTGCACACCAATGACTTTTTC 634
   ::::::::::: || ::::::::::: ||::| |::| |
63 eutHraSnuSerLeuLeuAspLeuGlyTyrThrGlySerThrValPro 79
   ::::::::::: || ::::::::::: ||::| |::| |
633 TCAGCCACCTGCTCTCTGCTGATCTTATTCACAGTGGTGGCT 584
   ::::::::::: || ::::::::::: ||::| |::| |
80 GlnMetLeuValAsnLeuGlySerIleArgLysValIleSerTyrArg 96
   ::::::::::: || ::::::::::: ||::| |::| |
583 AAGATGCTGTAGACCTACTGCTAGAAATACATCTATCTTTCTTGG 534
   ::::::::::: || ::::::::::: ||::| |::| |
96 yCysValAlaGlnLeuPheIlePheLeuAlaLeuGlyAlaThrGluTyr 113
   ::::::::::: || ::::::::::: ||::| |::| |
533 CAGTGTATGACAGTCTCTACCTTCTGATTTTATAGATGCGACTGTG 484
   ::::::::::: || ::::::::::: ||::| |::| |
113 euLeuLeuAlaValMetSerPheAspArgPheValAlaIleCysArgPro 129
   ::::::::::: || ::::::::::: ||::| |::| |
483 TCCTGCTGGCAGTGTATGATGATATAGCCATTTAGTATGATGATCCC 434
   ::::::::::: || ::::::::::: ||::| |::| |
130 LeuHsrIleSerValIleMetHsrGlnArgLeuGlyLeuLeuAlaI 146
   ::::::::::: || ::::::::::: ||::| |::| |
433 CTTTGTATGACAGTGTACATGTCCAGAAAGTGTCTTCATATATTGAC 384
   ::::::::::: || ::::::::::: ||::| |::| |
383 TGGTGTATTCTGTGGCATTTAGACAGATGCTTTGATACATACACATTGA 334
   ::::::::::: || ::::::::::: ||::| |::| |
163 hLeuGlnLeuProLeuGlyAspProTyrValIleAspHisPheLeuGly 179
   ::::::::::: || ::::::::::: ||::| |::| |
333 CATTCACATTATGTTCTGTGGTGCAGATGATATATCATTTCTTCTGT 284

```

```

180 GluValProAlaLeuLeuGlySerCysValGluThrThrAlaAsnG 196
   ::::::::::: || ::::::::::: ||::| |::| |
283 GATATCTCTCCAGTGTAGTACTGTCTGTTCAGACACAGGTCATATG 234
   ::::::::::: || ::::::::::: ||::| |::| |
196 uAlaGluLeuPheLeuValSerGluLeuPheHisLeuIleProLeuThr 213
   ::::::::::: || ::::::::::: ||::| |::| |
233 TTATGATCATATTCACGCTTTTGTTCATTTGATGACAGCCATCTCAG 184
   ::::::::::: || ::::::::::: ||::| |::| |
213 euLeuLeuSerTyrAlaPheIleValArgAlaValLeuArgIleGln 229
   ::::::::::: || ::::::::::: ||::| |::| |
183 GACTTCTATCTCTACTGTTACATCATCTCATCTGAGATCAGT 134
   ::::::::::: || ::::::::::: ||::| |::| |
230 SerAlaGluGlyArgGlnLysAlaPheGlyThrCysGlySerHisLeu 246
   ::::::::::: || ::::::::::: ||::| |::| |
133 TCAGCGCTGGAGGAGTTAAAGCTTCTCTACTGATACATCACACCTGAC 84
   ::::::::::: || ::::::::::: ||::| |::| |
246 eValValSerLeuPheTyrSerThrAlaValSerValTyrLeuGlnProp 263
   ::::::::::: || ::::::::::: ||::| |::| |
83 TGCTGTTCGCAATTTTCAAGGACTATGCTCTCATCATATTTTC.CGCCAA 35
   ::::::::::: || ::::::::::: ||::| |::| |
263 roSerProSerSerLysAspGlnGlyLysMet 273
   ::::::::::: || ::::::::::: ||::| |::| |
34 GTTCTTCTACTCTCTAGATCAAGATTAAGTG 3
   ::::::::::: || ::::::::::: ||::| |::| |
seq_name: gb_est2:BG197640

```

seq\_documentation\_block: 796 bp mRNA EST 21-APR-2001  
 LOCUS BG197640  
 DEFINITION RST17016 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG197640  
 VERSION BG197640.1 GI:13719455  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 796)  
 AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
 Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,  
 Lerner,L., Krishnoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,  
 Veloso,N., Hess,J., Cochren,K., Lo,K., Offenbacher,J., Danzig,J.  
 and Ducat,M.  
 TITLE Creation of Genome-wide Protein Expression Libraries using Random  
 Activation of Gene Expression  
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press  
 COMMENT Contact: Scott J. Cain  
 Athersys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596  
 Email: scain@atersys.com  
 FEATURES  
 source High quality sequence stop: 544.  
 location/Qualifiers  
 1..796  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Athersys RAGE Library"  
 /cell\_line="HT1080"  
 /note="See 'Creation of Genome-wide Protein Expression  
 Libraries using Random Activation of Gene Expression',  
 Nature Biotechnology, in press. Note that even though the  
 cell type indicated is HT1080, since a random activation  
 method was used, these sequence tags are not necessarily  
 expressed in HT1080 under normal circumstances."

alignment\_scores:  
 Quality: 490.50 Length: 252  
 Ratio: 2.666 Gaps: 2  
 Percent Similarity: 73.016 Percent Identity: 42.857









[illegible]

```
seq_name:/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-748-506-7
seq_documentation_block:
: Sequence 7, Application US/08748506
: Patent No. 6159707
: GENERAL INFORMATION:
: APPLICANT: Ronnelt et al.
: TITLE OF INVENTION: NOVEL SPERM RECEPTORS
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
STREET: Two Prudential Plaza, Suite 4500  
City: Chicago  
STATE: IL  
COUNTRY: US  
ZIP: 60601-6780  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/748,506  
FILING DATE: 08-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/033,751  
FILING DATE: 09-NOV-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
REFERENCE/DOCKET NUMBER: 74940  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5600  
TELEFAX: 312-616-5700  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 966 base pairs  
type: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-748-506-7

```
alignment_scores:
  Quality: 652.50      Length: 314
  Ratio: 2.825         Gaps: 2
  Percent Similarity: 73.567  Percent Identity: 42.994
```

```
alignment_block:
US-09-755-017-2 x US-08-748-506-7 .
```

Align seg 1/1 to: US-08-748-506-7 from: 1 to: 966

```

118 3 TrpValasnasperlle...ilegningupheilleleuLeuenglyPheSe 18
119 22 TGGCAGGAGATATGTTTGTCTGTCAACAGCCTTGATTCGCAAGTCTC 71
120 18 rAspArgProTrpLeuGluPheProLeuLeuValValPheLeuIleSerT 35
121 72 TGAAGTCCCTGGAGAAATGCTTCTCTCTGTTCACCCCTCATCTTCTCATGT 121
122 35 yTrhValThrIlePheGlyLeuLeuThrIleIleLeuValSerArgLeu 51
123 122 TCTTAGATCACACAAAGGAATACTACTCATATAGCCCTGCTATTGTGACC 171
124 52 AspThrLysLeuHisThrProMetLysrPhePheLeuThrAsnLeuSerle 68
125 172 AGTCATCTCTGTACACGCCCATGTACTCTTTCTGGCCAACTGTCTCT 221
126 68 uLeuAspLeuCysTyrrTrhThrCysrThrValProGlnMetLeuValAsnL 85
127 222 CTGGAGATTTGGGTATACTTGGCTGTGTATCCCAAGATGGTGGAGAGCC 271
128 85 euCysSerIleArgLysValIleSerTyrrArgGlyCysValAlaGlnLeu 101
129 272 TTGTGAGTGGAGGCCGAGGATCTCTTGGAGAGGTTGTGCTTCACAGATG 321
130 102 PheIlePheLeuAlaLeuGlyAlaThrGluTyrrLeuLeuLeuAlaValMe 118
131 322 TCTCTTCATCATATCTTTGTATTAAGTGAAGTGCCTGATTGGCAGGCAT 371
132 118 tSerPheAspArgPheValAlaIleCysArgProLeuHisTyrrSerValI 135

```

```

135 lemeHISGLArgLeuCySLeuGlnLeuAlaAlaSerIrrValThr 151
422 GAATAGTGTGGGGATGTGCTTATTTGGCAATTTGTCCATGGGGATG 471
152 GlyPheSerAsnSerValIrrPheSerThrLeuThrLeuGlnProle 168
472 GGATTCATAGTAGTGTGGGAGACACCAATTTATTTCTCTTGAACCT 521
168 uCysAspProTyrrValIleAspHisPheLeuCySLeuValProAlaLeu 185
522 CTGTGACCCCTGTGAGATAGACCACTCTTGTGTGACCTTCACCTGTC 571
185 euLysLeuSerCySValGluThrThrAlaAsnGlnAlaGluLeuPheLeu 201
572 TGGCACTTGGCTGTGTGATACATCCCAAAATGAGGCTCCATCTTTGTG 621
202 ValSerGluLeuPheHisLeuLeuProleThrLeuIleLeuSerTy 218
622 GCAGCAGTGTCTGCATATTAGTCCATTTTACTGATCATTTCTTCTTA 671
218 talapheIleValAlaValAlaValLeuArgIleGlnSerAlaGluGlyArg 235
672 TGTCAGAATTTCTGTTGGAGTGTGTGATGCTTCACCTGAGGGGCGCC 721
235 InLysAlaPheGlyThrCySgLySerHisLeuIleValValSerLeuPhe 251
722 ATAAAGCTCTCTCTACCTGTTCATCTCACCTGTGTATGTACACCTCTTC 771
252 TyrSerThrAlaValSerValIrrLeuGlnProProSerProSerSerTy 268
772 TATGGCTCAACATCTGCCACCTATTGTAGGTCCAAAGTCAAGCCACCTCC 821
268 sasplngLysMetValSerLeuPheTygIleIleAlaProMetL 285
822 AGGAGTGGACAAACHTTGGCCCTCTCTATATACATCAGTGCATCATCAGC 871
285 euAsnProLeuIleTyrrThrLeuArgAsnLysGluValLysGluGlyPhe 301
872 TGAATCCCATCTACATCAACGCTTAAGGACCAAGTAAGGATGAAGGCTGAC 921
302 LysArgLeuValAla.... ArgValPheLeuIleLysLys 313
922 ACAAGAACTCTGGGCTGAGAAAGTTCTGACAATGAAGAAG 963
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-467-948A-1
seq_documentation_block:
: Sequence 1, Application US/08467948A
: Patent No. 5998164
: GENERAL INFORMATION:
: APPLICANT: LI, YI
: APPLICANT: CAO, LIANG
: APPLICANT: NI, JIAN
: APPLICANT: GENTZ, REINER
: APPLICANT: BULT, CAROL J.
: APPLICANT: SUTTON III, GRANGER G.
: APPLICANT: ROSEN, CRAIG A.
: TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein
: TITLE OF INVENTION: Coupled Receptor GPR2
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVE., NW, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,948A
CLASSIFICATION: 435
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04079
FILING DATE: 30-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1140003/EKS/KLM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1713 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 116..1003
US-08-467-948A-1

```

alignment\_scores: Quality: 652.50 Length: 295  
Ratio: 2.900 Gaps: 3

Percent Similarity: 76.271 Percent Identity: 45.424

alignment\_block: US-09-755-017-2 x US-08-467-948A-1 ..

Align seg 1/1 to: US-08-467-948A-1 from: 1 to: 1713

```

5 AsnAspSerIleIleGlnLupheIleLeuLeuGlyPheSerAspArgP 21
1137 AATCAGACAAATGATGACAGATTCCTCTACCTGAGATTCCTCTGAGGCC 186
21 crrPheGlnLupheProLeuLeuValAlaPheLeuIleSerTyrrVal 38
187 AAGGATTCAGATGCTCTCTTGGGCTCTCTCCCTGTTCTATGTTCTCA 236
38 hrIlePheGlnLupheProLeuLeuValAlaPheLeuIleSerTyrrVal 54
237 CCTGCTGGGGAATGGGACCATCTGGGCTCATCTGACTGACATCCAGA 286
55 LeuHisThrProMetTyrrPhePheLeuThrAsnLeuSerLeuLeuAsp 71
287 CTCACACACCCCATGATGATCTTCTCTCTCACACCTGGCGGTGTAACAT 336
71 uCysTyrrThrCysThrValProGlnMetLeuValAsnLeuCySeri 88
337 CGCTATGGCTGCACACACAGTGGCCGACATGCGTGGGAACTCTCTCATC 386
88 leArgLysValIleSerTyrrArgGlyCysValAlaGlnLeuPheLeu 104
387 CAGCCAGCCCATCTCTCTTGGCTGTTGATGATGACATGATCTTCTTT 436
105 LeuAlaLeuGlyAlaThrGluTyrrLeuLeuAlaValMetSerPheAs 121
437 TTGAGTTTGCACATAGTGAATGCTGCTGTTGGTGTGATGCTGATGCA 486
121 parGheValAlaIleCysArgProLeuHisTyrrSerValIleMetHis 138
487 TCGGTAGTGGCCATCTGCCACCTCTCCGATATTTTCATCATCATGACCT 536
138 InArgLeuCySLeuGlnLeuAlaAlaSerIrrValThrGlyPheSer 154

```

```

537 GGAAGTCGATCAGTCTGGGATCAGTCTGGACATGTGGCTCCCTC 586
155 AsnSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeuGlySerAsp 171
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
587 CTGGCATGTGCATGTGAGCTCCTACCTTAAGACTGGCCCTTTGTGGCC 636
171 oTyValIleAspHisPheLeuGlySerGluValProAlaLeuLeuLysLeu 188
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
637 TCCTGAATCAACCACTCTTCTGTGAATCTGTCTGCTCCAGGCTGG 686
188 eTyValGluThrThrAlaAsnGlnAlaLeuLeuPheLeuValSerGlu 204
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
687 CCGTGTGTGACCTGCTGCTGACAGGTGCTGATCTTTGAAGCTGCATG 736
205 LeuPheHisLeuLeuProLeuThrLeuLeuLeuLeuLeuSerThrAlaPhe 221
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
737 TTGATCTGTGTGGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786
221 eValArgAlaValLeuArgIleGlnSerAlaGlnGlyArgGlnLysAla 238
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
787 CTGGGGGGGATCCTGAGATCCAGTCTGGGAGGGCCGAGAAAGGCT 836
238 heGlyThrGlySerHisLeuIleValValSerLeuPheThr 254
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
837 TCTCCACCTGCTCTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 886
255 AlaValSerValTrpLeuGlnProProSerProSerSerLysAspGln 271
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
887 GCATGCTCATGTACATGGCCCTAGTCCGCCATCTAGAGACGACGA 936
271 yLysMetVal.SerLeuPheTrpGlyIleLeuAlaPro..MetLysAsn 286
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
937 GAAGGTCCTTTCTATTATTACAGTCTTCAACCCGATGCTTAAAC 986
287 ProLeuIleThr.ThrLeuArgAsnLysGlu 296
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
987 CCCGTATTATTAACCCCTGAGGATGTAGAG 1017

```

## seq\_name: /cgn2\_6/plodata/2/lna/6A\_COMB.seq:us-08-467-947A-1

```

; seq_documentation_block:
; Sequence 1, Application US/08467947A
; Patent No. 6090575

```

```

; GENERAL INFORMATION:
; APPLICANT: LI, YI
; APPLICANT: CAO, LIANG
; APPLICANT: NI, JIAN
; APPLICANT: GENTZ, REINER
; APPLICANT: BULT, CAROL J.
; APPLICANT: SUTTON III, GRANGER G.
; APPLICANT: ROSEN, CRAIG A.
; TITLE OF INVENTION: Polynucleotides Encoding Human G-Protein
; TITLE OF INVENTION: Coupled Receptor GPR1
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PATENTIN RELEASE #1.0, VERSION #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,947A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04079
; FILING DATE: 30-MAR-1995

```

```

ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488,1140002/EKS/KLM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1713 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 116..1003
US-08-467-947A-1

```

```

alignment_scores:
Quality: 652.50 Length: 295
Ratio: 2.900 Gaps: 3
Percent Similarity: 76.271 Percent Identity: 45.424

```

## alignment\_block:

```

US-09-755-017-2 x US-08-467-947A-1 ..

```

```

Align seg 1/1 to: US-08-467-947A-1 from: 1 to: 1713

```

```

5 AsnAspSerIleIleGlnIlePheIleLeuLeuGlyPheSerAspArg 21
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
137 AATCAGACATGTGTACAGAGATTCCTGCTGAGATTTCTCCGGGCC 186
21 oTrpLeuGluPheProLeuLeuValValPheLeuIleSerThrVal 38
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
187 AAGCATTCAGATGCTCTCTTGGGCTCTTCCCTGCTCTGATGCTTCA 236
38 hTrpPheGlyAsnLeuThrIleIleLeuValSerArgLeuAspPhe 54
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
237 CCCTGCTGGGGAATGGACCATCTCGGGCTCATTCACAGCTCCAGCA 286
55 LeuIleThrProMetTrpPhePheLeuThrAsnLeuSerLeuAspLe 71
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
287 CTCACACCCCAATGATCTTCTCTGCTGACACTGGCCGCTGCAACAT 336
71 uCysTrpThrThrCysThrValProGlnMetLeuValAsnLeuCysSer 88
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
337 CGCTATGCTGCACACACTGCCCAAGATGCTGTGAACCTCTGCATC 386
88 LeuArgValIleSerTrpArgGlyCysValAlaGlnLeuPheIlePhe 104
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
387 CAGCCAGCCCATCTCTTGGTGTGATGACACTGACTTCTCTTT 436
105 LeuAlaLeuGlyAlaThrGluTrpLeuLeuLeuAlaValMetSerPhe 121
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
437 TTGAGTTTGCATACATGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 486
121 ArgPheValAlaIleCysArgProLeuHisTrpSerValIleMetHis 138
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
487 TCAGTATGTCGATGTCGACCCCTCCGATTTTCATCATCATGACCT 536
138 LnaArgLeuGlySerGlnLeuAlaAlaValSerTrpValTrpGlyPhe 154
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
537 GGAAGTCGATCAGTCTGGGATCAGTCTGGACATGTGGCTCCCTC 586
155 AsnSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeuGlySer 171
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
587 CTGGCATGTGCATGTGAGCTCCTACCTTAAGACTGGCCCTTTGTGGCC 636
171 oTyValIleAspHisPheLeuGlySerGluValProAlaLeuLeuLys 188
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
637 TCCTGAATCAACCACTCTTCTGTGAATCTGTCTGCTCCAGGCTGG 686

```



```

: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY:
: ZIP: 19406
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/827,291A
: FILING DATE: 28-MAR-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: King, William T
: REGISTRATION NUMBER: 30,954
: REFERENCE/DOCKET NUMBER: GP50001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-5015
: TELEFAX: 610-270-5090
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1290 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-827-291A-1.

alignment_scores:
      Quality: 648.00      Length: 301
      Ratio: 2.830      Gaps: 2
      Percent Similarity: 76.080      Percent Identity: 43.189

alignment_block:
US-09-755-017-2 x US-08-827-291A-1 ..
Align seg 1/1 to: US-08-827-291A-1 from: 1 to: 1290

5 AsnAspSerIleIleGlnGluPheIleLeuGly...PheSerAsp 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 AATTGACGCTTCACATCGACCTTCATCTCTCGGATCTTCATCAG 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
20 gPrOTPrleuGluPheProleuLeuValAlPheLeuIleSerTyThy 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
358 CCCGACCCACACCTCTCTCTCTCTGCTGCGCCATC...TTTTCAG 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37 alThrIlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAsp 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 TGCCCTTCATGGGAAACTCTGTGATGTTCTCTCACTTCACTGGACACC 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 LysLeuIleThrProMetTyRPhenPheLeuThrAsnLeuSerLeuLeu 70
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 CAGCTCCACACCCCATGTACCTCTCTCCACGCACTGCTCCTCAGGA 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 pleuCySTyRThrThCysThyValProGlnMetLeuValAsnLeuCys 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
505 CCTCATGCTCATCTGCACACCGTACCCCAAGATGGCTTCACTACCTGT 554
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 erIleArgLysValIleSerTyRArgGlyCysValAlaGlnLeuPheIle 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
555 CTGGGACGACAGTTCATTTCTATGCTGTGTGCCACACAAATTTTCTTC 604
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 PheLeuAlaLeuGlyAlaThrGluTyRLeuLeuAlaValMetSerPh 120

```

```

: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
605 TATACATCACTGCTTGCTGTAATGCTTCTTTGGCTGTATGGCTTA 654
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
120 eAspArgPheValAlaIleCysArgProLeuHisTyRSerValIleMet 137
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
655 TGACCCGCTACCTGCACTTTGCCACCTCTAAGACATCAATCTATGTA 704
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
137 IsGlnArgLeuCysLeuGlnLeuAlaAlaSerTrpValThrGlyPhe 153
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
705 GCCCTAAATTTGTGACCTTATGACGTGCTTTCCGATCCGCGGCTCT 754
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
154 SerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeuCys 170
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
755 ACAGATGGAATCATTTATGCTGTAGCCACATTTCTCTCCACGCTGG 804
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
170 pProTyRValIleAspHisPheLeuCysGlyValAlaProAlaLeuLys 187
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
805 GTCTGGGAAATAGCCCACTTCTCTGTGTGATGCTTCCCTCATATGTC 854
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
187 euserCysValGluThrThrAlaAsnGluAlaGluLeuPheLeuValSer 203
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
855 TCTCATGACATGACATCATATTTGAAAGSTATTTCATTTCTGCTCT 904
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
204 GluLeuPheHisLeuIleProLeuThrLeuIleLeuIleSerTyRAla 220
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
905 ATAGTATGCTGTTGTTCCCTGTTGCAATCATGCTCTCATATGCTGG 954
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
220 eIleValArgAlaValLeuArgGlyIleGlnSerAlaGluGlyArgGln 237
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
955 AGTTATCTGGCTGCTCATTCACATGGATCTGGAGAGGCTCGCGCAAG 1004
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
237 lApheGlyThrCysGlySerHisLeuIleValValSerLeuPheTyR 253
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
1005 CTTTCACAGACCTGCTCTCCTCACCCTCATGTGTGGAAAGTTCTATGA 1054
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
254 ThrAlaValSerValTyRLeuGlnProProSerProSerSerLysAsp 270
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
1055 GCAGGTTGTTTCATGTACATACAGCCACATCTGATCGCTCCCAACGA 1104
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
270 nGlyLysMetValSerLeuPheTyRArgGlyIleIleAlaProMetLeuAsp 287
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
1105 GGACAGCTGTGTGTATCTACACACATCCATCCATCCATCTCAATC 1154
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
287 roLeuIleTyRThrLeuArgAsnLysGluValLysGluGlyPheLysArg 303
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
1155 CCCTCATCTACAGCTCCGCAACAGAGAGTGCACAGACATTCATGAGAC 1204
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
304 Leu 304
: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
1205 ATC 1207

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:us-08-748-506-8
seq_documentation_block:
: Sequence 8, Application US/08748506
: Patent No. 6159707
: GENERAL INFORMATION:
: APPLICANT: Ronnett et al.
: TITLE OF INVENTION: NOVEL SPERM RECEPTORS
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Leydig, Volt & Mayer, Ltd.
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: IL
: COUNTRY: US
: ZIP: 60601-6780
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/748,506  
 FILING DATE: 08-NOV-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/033,751  
 FILING DATE: 09-NOV-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 REFERENCE NUMBER: 74940  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312-616-5600  
 TELEFAX: 312-616-5700  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 966 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-748-506-8

alignment\_scores:  
 Quality: 646.50 Length: 314  
 Ratio: 2.823 Gaps: 2  
 Percent Similarity: 72.930 Percent Identity: 42.994

alignment block:  
 US-09-755-017-2 x US-08-748-506-8 ..

Align seg 1/1 to: US-08-748-506-8 from: 1 to: 966

```

3 TrpValAsnAspSerIle...IleGlnIuPheIleuLeuGlyPhe 18
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
22 TGCGAGAGAAATGTTGTCTGTCAAAAGTTTGCAATTCACAAAGTTCTC 71
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
18 rAspArgProIleuGluPheProLeuLeuValaPheLeuIleSer 35
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
72 TGAGGTCCTCGAGAAATGCTCTCCGTCGTACCCCTCATCTCTCATGT 121
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
35 yThrValThrIlePheGlyAsnLeuThrIleLeuValaSerArgLeu 51
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
122 TCTTAGATCATACACAGCAATGCTCTCATAGCCCTTGCTATTGTAC 171
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
52 AspThrIleuHisThrProMetIlePhePheLeuThrAsnLeuSer 68
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
172 AGTCACATCTACACACCCCATGATCTCTTCTGCGCAACTGTGTCT 221
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
68 uLeuAspLeuCysTyrThrThrCysThrValProGlnMetLeuValAsn 85
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
222 CCGTGAGATGGGCTATACCTGCTGTGCATACCCCAAGATGCTCAGAGTC 271
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
85 euCysSerIleArgValIleSerTyrArgIleCysValaIleGlnLeu 101
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
272 TTGTGAGTGGGCCCGAGAGATCTTTCAGGTGGATGTGCCACACAGATG 321
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
102 PheIlePheLeuAlaLeuGlyAlaThrGluTyrLeuLeuLeuValaIle 118
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
322 TTTTCTTCATATCTTGTGTAATACGATGAGTGTCTGCTATTGTGAGCCAT 371
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
118 tSerPheAspArgPheValaIleLeuCysArgProLeuHisTyrSerVal 135
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
372 GCGCTTGAGCGGCTATATGCTATATGTCCCACTGCATATGCAACCC 421
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
135 leuMetHisGlnArgLeuCysLeuGlnLeuAlaIleAsnTyrValThr 151
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
422 GAATGAGTCTGAGATGTGCGCCACTTGGCAATTTGTCATGAGGTATG 471
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
152 GlyPheSerAsnSerValIlePheSerThrLeuThrLeuLeuProle 168
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
472 GGATGCAATAGTCTGGAGACAGCAAAATTTATTTTCTCTTGAAGCT 521
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
168 uCysAspProTyrValIleAspHisPheLeuCysGluValaProleuLeu 185

```

```

522 CTGTGAGACCCGTGATGATGACCACTCTCTGTGATCTCCACACTCTCC 571
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
185 euIleuSerCysValaGluThrThrAlaAsnGluAlaGluLeuPheLeu 201
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
572 TGGCATTGGCTGTGATGATACATCCCAATTTGAGGCTGCATCTTGTG 621
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
202 ValSerIleuPheHisLeuIlePheLeuThrLeuLeuLeuIleSerTyr 218
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
622 GTAGTGTCTCTGCAATATCTAGCCCTTTTGTGTCATTAATTTCTTA 671
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
218 rAlaPheIleValaArgIleGlnSerAlaGluIleArg 235
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
672 TGTGAGATTTCTGTGACAGTGTGATGATGCTTACACGAGGAGGCGCC 721
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
235 lnuAlaIlePheGlyThrCysGlySerHisLeuIleValaSerLeuPhe 251
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
722 ACAAGCCCTTTCACACTGTCTCCACCACTTGTATGATCAGCACTCTT 771
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
252 TyrSerThrAlaValSerValIleuGluProProSerProSerSerly 268
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
772 TATGCTAGAGATCTGTACCTATTTGAGGCTAAGCTAGCCACTAC 821
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
268 sAspGlnGlyLysMetValSerLeuPheTyrGlyIleIleAlaIleProMet 285
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
822 AGGATGAGCAAACTCTTGGCCCTCTCTACACAGCAGTACATCATCTGT 871
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
285 euAsnProLeuIleTyrThrLeuArgAsnLysGluValaLysGluIlePhe 301
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
872 TGAACCTATCATCTATATGTTAAGAAACAGATGTCAAGGCGCACTGT 921
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
302 LysArgLeuValaAla.....ArgValaPheLeuIleLysLys 313
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
922 AGAAGAAATCTGGCCCGAAAAAATTTCTGCAATTAATTAAG 963
  ||| :|||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:us-08-748-506-6
seq_documentation block:
Sequence 6, Application US/08/748506
Patent No. 6159707
GENERAL INFORMATION:
APPLICANT: Ronnett et al.
TITLE OF INVENTION: NOVEL SPERM RECEPTORS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voigt & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/748,506
FILING DATE: 08-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,751
FILING DATE: 09-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 74940
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5600
TELEFAX: 312-616-5700
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid

```

STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-748-506-6

alignment\_scores:  
 Quality: 636.50 Length: 302  
 Ratio: 2.893 Gaps: 1  
 Percent Similarity: 72.848 Percent Identity: 43.709

alignment\_block:

US-09-755-017-2 x US-08-748-506-6

Align seg 1/1 to: US-08-748-506-6 from: 1 to: 966

```

3 TrpValAsnAspSerile...ileglnlupheileuLeuLeuglyPhe 18
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
22 TGGCAGAGAGATAGTGTCTGTCACAAAGCTTTCATTTGCCAGCTTCTC 71
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
18 rAspArgProTrpLeuGlupheProleuLeuValValPheLeuIleSerT 35
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
72 TGAGGTCCCTGGAGAAATGCTTCCTGTCACCCCTCATCCTTCTCATGT 121
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
35 yTrThrValThrIlePheGlyAsnLeuThrIleIleLeuValSerArgLeu 51
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
122 TCTTAGATCACTACAGAGAAATCTCTCATAGCCCTTGCTGTTGTATCC 171
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
52 AspThrIysLeuHisThrProMetLysrPhePheLeuThrAsnLeuSerle 68
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
172 AGTCATCTCTACACACACCCCATGCTTCTTCTGCGCAACTGTCTCT 221
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
68 uLeuAspLeuLysrThrThrIleThrIleValProGlnMetLeuValAsnL 85
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
222 CCTGGAGATGGCTATAGTCTGCTGTCATACCCAGATGCTGCGAGGCC 271
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
85 eucysSerIleArgLysValIleSerTyraArgLysValAlaGlnLeu 101
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
272 TTGTGATGAGGCCAGAGAGATCTCTAGGGAGGATGTGCCACACATG 321
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
102 PheIlePheLeuAlaLeuGlyAlaThrGluTrpLeuLeuAlaValAlaMe 118
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
322 TTTTCTTTCACATTTTGGCATACATGAGTGGCTATTTGGCAGCAT 371
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
118 tSerPheAspArgPheValAlaIleCysArgProLeuHisTyrSerValI 135
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
372 GGCCTTGGACCGCTGATGGCATATGCTCCCATCTCCACTATGCAACC 421
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
135 leMetHisGlnArgLeuCysLeuGlnLeuAlaAlaIleSerTrpValThr 151
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
422 GAATGAGTCTGAGAGTATGTCCTTGGCAATTTGTTTCATGGGGAATG 471
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
152 GlyPheSerAsnSerValTrpLeuSerThrLeuThrLeuGlnLeuProle 168
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
472 GGATGCATAGTAGTCTGGAGACAGACCAATTNATTTNCTTGAAGCTT 521
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
168 uCysAspProTyraIleAspHisPheLeuGlyValAlaProAlaLeuL 185
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
522 CTGTGGACCTTGTGAGATAGACCACTTCTGTGACCTTCCACCTTCC 571
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
185 eulysLeuSerCysValGluThrThrAlaAsnGluAlaGluLeuPheLeu 201
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
572 TGGCAGCTGCTGTGATACATCCCAAAACAGAGCTGCCATCTTGTG 621
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
622 GCAGCAATCCTCTATATCTAGTCCATTTTGGTGATCTTATATCTTA 671
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
202 ValSerGluLeuPheHisLeuIleProLeuThrLeuLeuLeuIleSerT 218
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
622 GCAGCAATCCTCTATATCTAGTCCATTTTGGTGATCTTATATCTTA 671
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
218 rAlaPheIleValAlaValAlaLeuArgIleGlnSerAlaGluGlyArg 235
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
672 TGTGAGATTTCTGTTGATGCTGTGATGCTTACCTGAGGGGGGCC 721
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
235 lnySalapheGlyThrCysGlySerHisLeuIleValValSerLeuPhe 251
  
```

```

722 ATAAAGCTCTTCCACCTGTTCCACACCTACTACTAGTCACACTTTT 771
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
252 TySerThrAlaValSerValTyrrLeuGlnProProSerProSerSerly 268
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
772 TATGCTCTGTCTGCTTACCTATTGAGGCCCAAGCTAGCCACCTACC 821
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
268 sasPglngLysMetValSerLeuPheTyrglyIleIleAlaProMetL 285
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
822 AGCAATGACAAACTTGGCCCTTCTACACAGCAGATGATCAATCCATGC 871
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
285 euAsnProLeuIleTyrrThrLeuArgAsnLysGluValLysGluGlyPhe 301
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
872 TGAAACCTATCATCTACAGTCTAGAGACAGACAGATCAAGCAGCACTG 921
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
302 LysArg 303
  ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
922 AGAAGA 927
  
```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-748-506-9

seq\_documentation\_block:

Sequence 9, Application US/08748506

Patent No. 6159707

GENERAL INFORMATION:

APPLICANT: Ronnett et al.

TITLE OF INVENTION: NOVEL SPERM RECEPTORS

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: Leydig, Volt & Mayer, Ltd.

STREET: Two Prudential Plaza, Suite 4900

CITY: Chicago

STATE: IL

COUNTRY: US

ZIP: 60601-6780

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/748,506

FILING DATE: 08-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/033,751

FILING DATE: 09-NOV-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

REFERENCE/DOCKET NUMBER: 74940

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-616-5600

TELEFAX: 312-616-5700

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 984 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-748-506-9

alignment\_scores:

Quality: 587.00 Length: 319  
 Ratio: 2.552 Gaps: 3  
 Percent Similarity: 72.100 Percent Identity: 36.364

alignment\_block:

US-09-755-017-2 x US-08-748-506-9

Align seg 1/1 to: US-08-748-506-9 from: 1 to: 984





```

28 uValValPheLeuIleSerThrValThrIlePheGlyAsnLeuThrI 45
   ::::::::::::::::::::
864 TGGGCTGTCTCTGCATGCTGCTGACGGGTGGGGAACCTGCTCA 913
   ::::::::::::::::::::
45 IleIleuValSerArgLeuAspThrLysLeuHisThrPrometLysPhe 61
   ::::::::::::::::::::
914 TCATCCGCGGCATCAGCCCTGACTCCACCTCCACACCCCATGTAATTC 963
   ::::::::::::::::::::
62 PheLeuThrAsnIleuSerLeuLeuAspLeuLysThrThrLysThrVal 78
   ::::::::::::::::::::
964 TTCTCTCCCAACCTGTCTGCTGCTGACATCGGTTTCACCTCCACACGGT 1013
   ::::::::::::::::::::
78 1ProGlnMetLeuValAsnLeuLysSerLeuArgLysValIleSerVal 95
   ::::::::::::::::::::
1014 CCCCAAGATGATGTGTGACATCCAGCTCAGACAGAGATGATCTCCATG 1063
   ::::::::::::::::::::
95 rglGlyValAlaGlnLeuPheIlePheLeuAlaGlyAla...Thr 110
   ::::::::::::::::::::
1064 CAGGCTGCCGACATCAGATGCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1113
   ::::::::::::::::::::
111 GluThr LeuLeuLeuAlaValMetSerPheAspArgPheValAlaIleC 127
   ::::::::::::::::::::
1114 GAGAGACATCTCTCTGATGATGAGCTGATGAGCTGATGAGCTGATGAG 1163
   ::::::::::::::::::::
127 ySarArgProLeuHisLysSerValIleMetHisGlnArgLeuLysGln 143
   ::::::::::::::::::::
1164 GTCAACCGCGATATGATTCACAC.ATCATGAAACCGGTGTTCTGTGCTT 1212
   ::::::::::::::::::::
144 LeuAlaAlaLaser...TrpValThrGlyPheSerAsnSerValTrrple 159
   ::::::::::::::::::::
1213 CTAGATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1262
   ::::::::::::::::::::
159 userThrLeuThrLeuGlnLeuProLeuLysAspProTyrValIleAspH 176
   ::::::::::::::::::::
1263 CAACCTATGTCCTTACACAGTGCCTTCCAGAGATGCGAATTCCTA 1312
   ::::::::::::::::::::
176 1sPheLeuLysGlnValProAlaLeuLeuLysLeuSerCysValGluThr 192
   ::::::::::::::::::::
1313 ATTTCTCTGTGACCTTCTTCACTCTCCCATCTTGCATGTGTGACACC 1362
   ::::::::::::::::::::
193 ThrAlaAsnGlnAlaGlnLeuPheLeuValSerGlnLeuPheHisLeuI 209
   ::::::::::::::::::::
1363 TTCACCATTTAATCATGATGATTTCCCTGCTGCTGCTGCTGCTGCTT 1412
   ::::::::::::::::::::
209 eProLeuThrLeuIleLeuIleSerThrAlaPheIleValArgAlaVal 226
   ::::::::::::::::::::
1413 TCCCATCTCAGGACCTTCTCTTACTGTAATAATTCCTTCCCTCCTTC 1462
   ::::::::::::::::::::
226 euArgIleGlnSerAlaGlnGlyArgGlnLysAlaPheGlyThrCysGly 242
   ::::::::::::::::::::
1463 TGAAGGTTTCATCATCAGTGGGGAATTAACCT.TTCTCCACCTGTGGG 1511
   ::::::::::::::::::::
243 SerHisLeuIleValVal.SerLeuPheTyrSerThrAlaValSerValT 259
   ::::::::::::::::::::
1512 TTCACCTCTCAGTGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1540
   ::::::::::::::::::::
259 yTrpLeuGlnProProSerProSerSerLysAspGlnGlyLysMetValSer 275
   ::::::::::::::::::::
1541 ACCTCGGTAGATGTGTCTCTTCCCGGAGAAAGAGTGCAGTGGCTCA 1590
   ::::::::::::::::::::
276 LeuPheThrGlyIleIleAlaPrometLeuAsnProLeuIleTyrThrLe 292
   ::::::::::::::::::::
1591 GTGATGTACACGGGTGCTCACCCTCATGACCCCTTCATGTATGACACCT 1640
   ::::::::::::::::::::
292 uArgAsnLysGlnValLysGlnGlyPheLysArg 303
   ::::::::::::::::::::
1641 GAGAAACAGGATATGAAAGTCTCTGCGCGCG 1674
   ::::::::::::::::::::
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-465-980-1
seq_documentation_block:
; Sequence 1, Application US/08465980
; Patent No. 5756309

```

```

GENERAL INFORMATION:
APPLICANT: Soppet, Daniel R.
APPLICANT: Li, Yi
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPAJ70
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
ADDRESSER: STUART & OLSTEIN
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,980
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-446
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1474 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 274..1233
US-08-465-980-1

alignment_scores:
Quality: 385.00 Length: 279
Ratio: 2.163 Gaps: 2
Percent Similarity: 63.799 Percent Identity: 31.900

alignment_block:
US-09-755-017-2 x US-08-465-980-1 ..

Align seq 1/1 to: US-08-465-980-1 from: 1 to: 1474

22 TrpLeuGlnPheProLeuLeuValAlaPheLeuIleSerThrValAlaThr 38
   ::::::::::::::::::::
346 TGGGTTGGCTTCCCTCTTCCATG.....TATGTAAGGCG 383
   ::::::::::::::::::::
38 rIlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLysL 55
   ::::::::::::::::::::
384 AATCTGTGCAAACTGCATCTGCTTTCATCGTAAGACGACGACGACGAC 433
   ::::::::::::::::::::
55 euHisThrPrometLysPhePheLeuThrAsnLeuSerLeuLeuAspLeu 71
   ::::::::::::::::::::
434 TGCACGCTCGATGATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
   ::::::::::::::::::::
72 CysThrThrThrCysThrValProGlnMetLeuValAsnLeuLysSerI 88
   ::::::::::::::::::::
484 GCCTTATCCACATCCACATGCTTATGATGCTTCTGCTGCTTCTGCTTGA 533
   ::::::::::::::::::::
88 eArgLysValIleSerThrArgGlyCysValAlaGlnLeuPheIlePheL 105
   ::::::::::::::::::::
534 TTCCCGAGATTAAGATTGAGGCTGTCTTACCCAGATGTCTTATATTC 583
   ::::::::::::::::::::

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/053,303
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/465,980
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-446
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1474 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 274..1233
US-09-053-303-1

alignment_scores:
Quality: 385.00 Length: 279
Ratio: 2.165 Gaps: 2
Percent Similarity: 63.799 Percent Identity: 31.900

alignment_block:
US-09-755-017-2 x US-09-053-303-1 ..

Align seg 1/1 to: US-09-053-303-1 from: 1 to: 1474

22 TrpLeuGluPheProLeuLeuValValPheLeuLeuSerThrValTh 38
||||| ||||||||| ||| |||||||
346 TGGGTGGCTTCCCTCTCTTCATG.....TATGTAGTGGC 383
38 rIlePheGlyLeuThrIleIleLeuValSerArgLeuAspThrLysL 55
||| ||||| ||||||||| ||| ||| |||
384 AATGTGTGAAGTGCATCGTGGTCTTCATCGTAAGACGGAAGCAGCC 433
55 euHisThrProMetLysrPhePheLeuThrAsnLeuSerLeuAspLeu 71
||||| ||||||||| ||||||| ||| |||
434 TGCAGCGCTCCCATGTACCTCTTCTGCAATGCTTGAGCCATGGACCTG 483
72 CysThrThrThrCysThrValProGlnMetLeuValAsnLeuCysSerI 88
||| ||||| ||||||||| ||||||| |||
484 GCCTTATCCAAATCCACCATGCTTAAGATCTTCCCTTTTGCTTGA 533
88 eaArgValIleSerThrArgGlyCysValAlaGlnLeuPheIlePheL 105
||| ||||| ||||||||| ||||||| ||| |||
534 TTCCCGAAGATGTAGCATGTAGGCGTGTATACCAAGATGTCTTATTTC 583
105 euAlaLeuGlyAlaThrGluThrLeuLeuLeuAlaValMetSerPheAsp 121
||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
584 ATGCCCTCTCCCATGTAAATCCACACATCCGCTGGCCATGAGCCTTTGAC 633
122 ArgPheValAlaIleCysArgProLeuHisThrSerValIleMetHisG 138
||||| ||||||||| ||||||||| ||||||| ||| ||| ||| |||
634 CGTATGTGGCCATGTGCACCCATGCGGCGATCTGCAGAGCTCAACAA 663
138 nArgLeuLysLeuGlnLeuAlaAlaAlaSerLysThrValThrGlyPheSer 155
||| ||||| ||||||||| ||||||| ||| ||| ||| ||| |||
684 TACAGTACACAGCCAGATGTGCATCGTGGCGTGGTCCGGGAGATCCCTCT 733
155 snSerValThrLeuSerThrLeuThrLeuGlnLeuProLeuCysAspPro 171
||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
734 TTTTTCCTCCATGCTCTCTGTATCAAGCGGCTGGCGCTTCTGCACTCC 783

```

```

172 TyrValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeu 188
    |||:|||||
784 AATGTCCTCGCACCTCTATGTGTGTCCACAGATGATATGAACTGGC 833
    |||:|||||
188 rcysValGluThrThralaasnGluValLeuPheLeuValSerGlu 205
    :|||||:|
834 CTATGCAGACCTTGGCCAAATGTGTATGTGCTTACTGCTTCCATCTGC 883
    |||:|||||
205 eupheHisLeuIleProLeuThrLeuLeuLeuSerTyrAlaPheIle 221
    |||:|||||
884 TGCTCATGGCGGTGGACGTAATGTCTCTTGTCTTATTTTCTGATA 933
    |||:|||||
222 ValArgAlaValLeuArgIleGlnSerAlaGluGlyArgGlnLysAla 238
    :|||||:|||||
934 ATACGAAGGCTTGCACACTGCTCTCCAACTCAGAGCGGCGCAAGGCTT 983
    |||:|||||
238 eGlyThrCysGlySerHisLeuIleValValSerLeuPheTyrSerThra 255
    |||:|||||
984 TGGACCTGTGTGTGCACACATGTGTGTGTGCTGCGCTTATGTGCCAC 1033
    |||:|||||
255 ValAlaSerValTyrLeuGlnProProSerProSerSerLysAspGlnGly 271
    |||:|||||
1034 TTATGTGCTCTCAGTTGTACACCGCTTTGGAACACCTTCATCCCAT 1083
    |||:|||||
272 LysMetValSerLeu.....PheTyrGlyIleIleAlaProMetLeuAs 286
    |||:|||||
1084 GTCCGTGTGTGTGTGATGCTGATCTGCTGCTGCTGCTGCTGCTGCTG 1133
    |||:|||||
286 nProLeuIleTyrThrLeuArgAsnLysGluValLys 298
    |||:|||||
1134 TCCCATCATCTATGTGTGCCCAAAACCAAAACAGATCAGA 1170

```

seq\_name: /sgn2\_6/ptodata/2/1na/PCrUS\_COMB.seq: PCT-US95-07093-1

seq\_documentation\_block:

Sequence 1, Application: PC/TUS9507093

GENERAL INFORMATION:

APPLICANT: Soppet, Daniel R.

APPLICANT: Li, Yi

APPLICANT: Rosen, Craig A.

APPLICANT: Ruben, Steven M.

TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPRJ70

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,

ADDRESSER: STUART & OLSTEIN

STREET: 6 Becker Farm Road

CITY: Roseland

STATE: New Jersey

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/07093

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/465,980

FILING DATE: 06-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Ferraro, Gregory D.

REGISTRATION NUMBER: 36,134

REFERENCE/DOCKET NUMBER: 325800-446

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

```

? LENGTH: 1474 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 274..1233
? PCT-US95-07093-1

alignment_scores:
    Quality: 385.00    Length: 279
    Ratio: 2.163      Gaps: 2
    Percent Similarity: 63.799    Percent Identity: 31.900

alignment_block:
US-09-755-017-2 x PCT-US95-07093-1 ..
Align seg 1/1 to: PCT-US95-07093-1 from: 1 to: 1474

22 TrpLeuGlnPheProLeuLeuValAlaPheLeuIleSerTyrThrValTh 38
    |||:|||||
346 TGGGTGGCTTCCCTCTCTTCCATG.....TATGTAGTGGC 383
    |||:|||||
38 rIlePheGlyAsnLeuThrIleIleLeuValSerArgLeuAspThrLysL 55
    :|||||:
384 AATGTGTGGAACTGCATGCTGTGTCTTCATCTGTAAGGACGAGACGACGC 433
    |||:|||||
55 euhIsthrProMetTyrPhePheLeuThrAsnLeuSerLeuLeuAspLeu 71
    |||:|||||
434 TGCACGCTCGATGTACCTCTTCTGTGCATGCTTGCACGACCATGTGACC 483
    |||:|||||
72 CysTyrThrThrCysThrValProGlnMetLeuValAsnLeuGlySerIle 88
    |||:|||||
484 GCCTTATCCACATCCACATCCATCCATCCATCCATCCATCCATCCATCC 533
    |||:|||||
88 eArgLysValIleSerTyrArgGlyCysValAlaGlnLeuPheIlePheIle 105
    :|||||:
534 TTCCCGAGAGATTAAGCATTAAGGCGCTGTCTTACCAGATGTTCTTATTC 583
    |||:|||||
105 euhAlaLeuGlyAlaThrGluTyrLeuLeuLeuAlaValMetSerPheasp 121
    |||:|||||
584 ATCCCTCTCAGCATTAATCCACCATCTCTGCTGCGCATGCGCTTGTGAC 633
    |||:|||||
122 ArgPheValAlaIleCysArgProLeuHisTyrSerValIleMetHisGlu 138
    |||:|||||
634 CGTTATGTGGCCATGTGCCACCCACGCGGCATGTGCATGTGCACAA 683
    |||:|||||
138 nArgLeuCysLeuGlnLeuAlaAlaIleSerTrpValIleThrGlyPheSer 155
    |||:|||||
684 TACAGTAACAGCCACAGATTGGCATGCTGGCTGTGGTCCGGGATCCCTCT 733
    |||:|||||
155 snSerValTrpLeuSerThrLeuThrLeuGlnLeuProLeuCysAspPro 171
    |||:|||||
734 TTTTTCACACGCTCTGCTGTATGATCAAGGCGCTTCTGCACCTCC 783
    |||:|||||
172 TyrValIleAspHisPheLeuCysGluValProAlaLeuLeuLysLeu 188
    |||:|||||
784 AATGTCCTCGCACCTCTATGTGTGTCCACAGATGATATGAACTGGC 833
    |||:|||||
188 rcysValGluThrThralaasnGluValLeuPheLeuValSerGlu 205
    |||:|||||
834 CTATGCAGACCTTGGCCAAATGTGTATGTGCTTACTGCTTCCATCTGC 883
    |||:|||||
205 eupheHisLeuIleProLeuThrLeuLeuLeuSerTyrAlaPheIle 221
    |||:|||||
884 TGCTCATGGCGGTGGACGTAATGTCTCTTGTCTTATTTTCTGATA 933
    |||:|||||
222 ValArgAlaValLeuArgIleGlnSerAlaGluGlyArgGlnLysAla 238
    |||:|||||
934 ATACGAAGGCTTGCACACTGCTCTCCAACTCAGAGCGGCGCAAGGCTT 983
    |||:|||||

```



alignment\_block:  
US-09-755-017-2 x US-08-436-074-57/rev ..

Align seg 1/1 to reverse of: US-08-436-074-57 from: 1 to: 1320

```
240 ThrCysGlySerHisLeuIleValSerLeuPheTyrSerThrAlaVal 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1319 ACTTGAGCTCGCATCTAATTGTTGCTCTCTCTATAGTACAAATCAT 1270
256 lSerValTyrLeuGlnProProSerProSerSerLysAspGlnGlyLysM 273
: |||||:|||||: |||||: |||||: |||||: |||||: |||||:
1269 ATACATGTACTCCAGCCAGCAAAATCTTATTCCTCCAGACCGGCAAGT 1220
273 eValSerLeuPheTyrGlyIleIleAlaProMetLeuAsnProLeuIle 289
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
1219 TTCTTACCCCTTTCTACACAAATTGTCACTCCAGTGTTAACCCCTGATC 1170
290 TyrThrLeuArgAsnLysGluValLysGluGlyPheLysArgLeuValAl 306
|||||:|||||: |||||: |||||: |||||: |||||: |||||:
1169 TATACACTAAGAAACAAGATGTTAAAGAGGCATGAAAGAGGTGCTAGG 1120
306 aArg 307
: |||||:
1119 GAAG 1116
```